Investigating the interactions of G-quadruplexes with transcription and chromatin structure

Jiazhen Shen

Cancer Research UK Cambridge Institute

St. Catharine's College

University of Cambridge

This dissertation is submitted for the degree of Doctor of Philosophy

February 2021

Declaration

This thesis is the result of my own work. Bioinformatic analysis of sequencing data is carried out in collaboration with Dr. Angela Simeone. The in-house synthesis of pyPDS was performed by Dr. Xiaoyun Zhang and CD spectra were obtained by Dr. Shiqing Mao. No other work was carried out in collaboration except as specified in the text. I state that my dissertation is not substantially the same as any work that I have submitted before, or, is being concurrently submitted for any degree, diploma or other qualification at the University of Cambridge or any other University or similar institution unless explicitly stated in the text. In accordance with the guidelines of the Degree Committee for the Faculties of Clinical Medicine and Veterinary Medicine, this thesis does not exceed the word limit of 60,000 words.

Jiazhen Shen
St. Catharine's College
Cambridge
January 2021

Summary

Investigating the interactions of G-quadruplexes with transcription and chromatin structure

Jiazhen Shen

Four stranded G-quadruplexes (G4s) are non-canonical nucleic acid secondary structures that are primarily found in nucleosome-depleted promoters and mark sites of elevated transcription within the human genome. Although numerous studies have indicated an association between G4 formation and active transcription, the causal nature of this relationship remains unknown. The work in this thesis first addresses whether promoter G4s form as a result of transcription or whether G4 formation enhances transcription. In addition, this thesis queries the relationship between the folding of G4s and chromatin accessibility at promoters.

The dependence of promoter G4 formation on transcription was tested following Inhibition of transcription elongation and initiation by 5,6-dichlorobenzimidazole 1-ß-D-ribofuranoside and triptolide respectively. Using G4 ChIP-seq, it was found that promoter G4 formation was not diminished upon transcriptional inhibition, refuting the dogma that folding of promoter G4s depends on transcriptional activity. Since endogenous G4s form primarily within sites of accessible chromatin, the impact of rapid chromatin compaction on promoter G4 folding was then assessed. Chromatin compaction was found to cause a depletion of some promoter G4s (~20%), which accompanied a reduction in RNA polymerase II (RNA Pol II) loading. This indicates that chromatin accessibility, as opposed to transcriptional activity, is a determinant of G4 formation. To assess whether G4s function to promote transcription, the small molecule ligand pyPDS was used to stabilise G4s in compacted chromatin. This was found to diminish the loss of G4s and mitigated loss of RNA Pol II at promoters upon chromatin compaction, indicating a role of G4s in promoting polymerase occupancy.

Overall, these results show that G4 folding within promoters is fostered by the establishment of accessible chromatin and functions to enhance transcription by promoting Pol II loading. This builds on earlier work showing that presence of promoter G4s accompanies increased gene expression and provides evidence that the G4 structure itself may be sufficient to drive transcription by recruiting key components of the transcriptional machinery.

Acknowledgements

First of all, I would like to thank my PhD supervisor Professor Sir Shankar Balasubramanian for providing me such an excellent opportunity to carry out my PhD projects with great experimental freedom. I am grateful for all his invaluable support, encouragement, guidance and generous advice throughout my study. I feel very lucky to be one of the members in his lab. I would also like to express my gratitude to Dr. David Tannahill, for his daily support and suggestions. It would be very hard for me to complete my PhD without his guidance.

In addition, I would like to express my deepest appreciation to all the amazing people I have worked with during my study. I thank Dr. Dhaval Varshney for his patience, supervision and constructive suggestions; Dr. Angela Simeone for the data analysis she's done and for her inspired discussions; Dr. Katie Zyner, Dr. Jochen Spiegel, Dr. Robert Hansel-Hertsch and Dr. Eun-Ang Raiber-Moreau for their assistance with protocols, experiments and troubleshooting; Dr. Sean Flynn for his generous advice on the scientific writing skills. I also want to thank other members of the Balasubramanian group past and present, notably Shiqing, Yuqi, Xiaoyun, Angie, Zutao, Silvia, Winnie, Areeb, Darcie, Kim, Jane, Antanas, Pradnya, Isabel, Matt, Santosh, Jo, Marco, Guillem and Sergio. I consider many of you my true friends and thank you for creating such a supportive and amazing environment, and for providing stimulating discussions. Thank you for all your help and making my life as a PhD student enjoyable and fun.

I wish to acknowledge Cancer Research UK for their funding which offers the freedom and flexibility for my research. I am grateful to the CRUK-CI core facilities, in particular the Genomics core, Light Microscopy core and Research Instrumentation & Cell Services for help in this work. I also want to thank the Scientific Administration team, especially Dr. Ann Kaminski and Mrs Julie Housden, for all their help.

Finally, I will never forget my friends and family, Mum and Dad, my grandparents and Shiqing, for their understanding and patience. Thank you for being a constant source of love, support and encouragement throughout my PhD study and my life.

Contributions to publications

Di Antonio M., Ponjavic A., Radzevičius A., Ranasinghe R.T., Catalano M., Zhang X., **Shen J.**, Needham L.M., Lee S.F., Klenerman D., Balasubramanian S. Single-molecule visualization of DNA G-quadruplex formation in live cells. *Nat Chem* **12**, 832-837 (2020).

My contribution to the above manuscript involved studying the dynamic effect of dimethyl sulfate (DMS)-mediated methylation on G4 structures. These data are not discussed in the thesis.

Shen J., Varshney D., Simeone A., Zhang X., Adhikari S., Tannahill D. and Balasubramanian S. Promoter G-quadruplex folding precedes transcription and is controlled by chromatin. *Genome Biol* **22**, 143 (2021).

At the time of submission of this thesis, the majority of this work described in this thesis has been accepted and published in Genome Biology.

Contents

Table of Figures

Abbreviations

8-oxoG	8-oxo-7,8-dihydroguanine
APE1	Apurinic/apyrimidinic endonuclease 1
ATAC-seq	Assay for transposase-accessible chromatin using sequencing
BER	Base excision repair
bp	Base pair
BSA	Bovine serum albumin
CD	Circular dichroism
ChIP-seq	Chromatin immunoprecipitation followed by high-throughput sequencing
CNA	Copy number alteration
cPDS	Carboxypyridostatin
CPSF	Cleavage and polyadenylation specificity factor
CstF	Cleavage stimulatory factor
CTCF	CCCTC-binding factor
CTD	Carboxy-terminal domain
DMEM	Dulbecco's Modified Eagle Medium
DMS	Dimethyl sulfate
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DRB	5,6-dichlorobenzimidazole 1-ß-D-ribofuranoside
DSIF	DRB-sensitivity-inducing factor (DSIF), negative elongation factor
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
eIF4A	Eukaryotic initiation factor 4A
ELISA	Enzyme-linked immunosorbent assay
FBS	Fetal bovine serum
FLIM	Fluorescence lifetime imaging microscopy
FRET	Förster resonance energy transfer
FUSE	Far upstream element
G	Guanine
G4	Guanine-quadruplex

G4 ChIP-seq	G4 chromatin immunoprecipitation and high-throughput sequencing
G4-seq	G4-sequencing
H3K4me3	Trimethylation of H3 at lysine 4
H3K9me3	Trimethylated lysine 9 of histone H3
H3K56	Lysine 56 in histone H3
H4K20	Lysine 20 of H4
HAT	Histone acetyltransferase
HDAC	Histone deacetylase
HIF1α	Hypoxia-inducible factor-1α
HP	Heterochromatin protein
HR	Homologous recombination
IF	Immunofluorescence
ILPR	Insulin linked polymorphism region
LM-PCR	Ligation-mediated polymerase chain reaction
MAZ	Myc-associated zinc finger
MNase	Micrococcal nuclease
MOPS	3-(N-morpholino) propanesulfonic acid
mtDNA	Mitochondrial DNA
MW	Molecular weight
NAI	2-methylnicotinic acid imidazolide
DNR	Nucleosome depleted region
NELF	Negative elongation factor
NHEJ	Nonhomologous end joining
NMR	Nuclear magnetic resonance
nt	nucleotide
OQS	Observed G4 structures
P-TEFb	Positive transcription elongation factor b
PAGE	Polyacrylamide gel electrophoresis
PAFc	Pol-II-associated factor complex
PARP-1	Poly(ADP-ribose) polymerase 1
PBS	Phosphate-buffered saline
PDAC	Pancreatic ductal adenocarcinoma

PDS	Pyridostatin
PIC	Protein inhibitor cocktail
PQS	Putative quadruplex sequence
pyPDS	Pyrrolidine pyridostatin
qPCR	Quantitative PCR
RIP	RNA immunoprecipitation
RIPA	Radioimmunoprecipitation assay buffer
RNA	Ribonucleic acid
RNA Pol II	RNA polymerase II
ROI	Regions of interest
RT	Room temperature
scFv	Single-chain variable fragment
SD	Standard deviation
SDS	Sodium dodecyl sulphate
ssDNA	Single-stranded DNA
TAD	Topologically associated domain
TBS	Tris-buffered saline
TEBP	Telomere end-binding protein
TERRA	Telomeric repeat-containing RNA
TF	Transcription factor
tiRNA	tRNA-derived stress-induced RNAs
TMB	3,3’,5,5’-tetramethylbenzidine
TPL	Triptolide
TSS	Transcription start site
UTR	Untranslated region
UV	Ultraviolet
VEGF	Vascular endothelial growth factor

Chapter 1

General introduction

The work described in this thesis focuses on the regulation of non-canonical DNA secondary structures known as guanine-quadruplexes (G4s), which have been implicated in various cellular functions. DNA G4s are particularly prevalent at promoters and accumulating evidence implicates them in transcriptional regulation. At the onset of my studies, however, the mechanism underlying the spatiotemporal regulation of G4 formation in transcriptional regulatory regions was largely unknown. My primary aim is to explore the functional relationship between the folding of G4s, transcription and chromatin state.

The experimental work is presented in three chapters. In Chapter 2, I focus on studying the causal relationship between G4 folding and transcription. Here, I also validate tools and techniques used throughout the study. In Chapter 3, I investigate the interplay between the folding of G4 structures and accessible chromatin state. I next explore mechanisms by which G4s may impact transcription in Chapter 4. My overall findings are summarised in Chapter 5, where I discuss their implications for our understanding of the functions of G4s in transcription and chromatin state.

The main areas of interest in this thesis are introduced in the first chapter, which provides an overview of G4s, including their chemistry and biology, as well as tools used to study G4s *in vitro* and in cells. To provide context to models that propose interplay between G4s and transcription, I also overview key principles of transcription machinery. Finally, I describe unanswered questions that remain in the field, and how my PhD project aims to address them.

1.1 G-quadruplexes: from sequences to structures

1.1.1 G-tetrads and G-quadruplexes

In 1953, Rosalind Franklin and Maurice Wilkins reported high quality X-ray diffraction images of DNA fibers (Franklin and Gosling, 1953; Wilkins et al., 1953). Franklin's experimental rigour led to the first conceptual model of double-stranded DNA helix described by James Watson and Francis Crick, which emerged in the same year (Watson and Crick, 1953). This helical structure of DNA, which is maintained by Watson-Crick hydrogen bonding, has been shown to be the predominant form in eukaryotic cells and is also known as canonical B-DNA (Richmond et al., 1986). However, accumulating evidence suggests that DNA is structurally dynamic. Studies have shown that the interactions between nucleotides can adopt alternative hydrogen bonds to classic Watson-Crick ones to form non-canonical secondary structures, such as H-DNA, Z-DNA, cruciforms and guanine(G)-quadruplexes (Bochman et al., 2012; Murat and Balasubramanian, 2014) (Figure 1.1). In this dissertation, I will focus on one of these non-canonical structures known as G-quadruplexes (G4s).

In 1910, Bang reported that concentrated guanylic acid formed a gel-like viscous solution (Bang, 1910). This unusual physical property was not further investigated until 1962, when X-ray fiber diffraction patterns illustrated four-fold helix structures that had further been suggested as planar tetramolecular structures (Gellert et al., 1962). Further evidence suggesting the existence of G4s showed that synthetic guanine (G)-rich DNA derived from the human genome (Sen and Gilbert, 1998) and from Oxytricha telomeric sequence (Williamson et al., 1989) formed secondary structures and caused reduced electrophoretic mobility. In 1990s, the presence of RNA G4s were subsequently reported *in vitro* (Christiansen et al., 1994; Kim et al., 1991).

Rather than interacting with cytosine via Watson-Crick hydrogen bonds (Figure 1.2A), guanine residues in G-rich sequences act as both donor and acceptor of Hoogsteen hydrogen bonds and associate with their adjacent Gs to form a planar tetrameric quartet, known as G-quartet or G-tetrad (Figure 1.2B). A monovalent cation in the quartet core provides further stability to the G-tetrad by neutralising the electrostatic repulsion from O_6 of each guanine, with the order of stability $K^+ > Rb^+ > Na^+ > Cs^+ > Li^+$ (Baumann, 2010). Self-

stacking of multiple G-tetrads via π-π interactions further stabilises the structure and forms a G4, which is linked by loops of nucleotides (Figure 1.2B). Although a stable G4 structure usually consists of three or more layers of G-tetrads, two-G-tetrad conformations with high stability have also been reported in DNA (Lim et al., 2009; Zhang et al., 2010b).

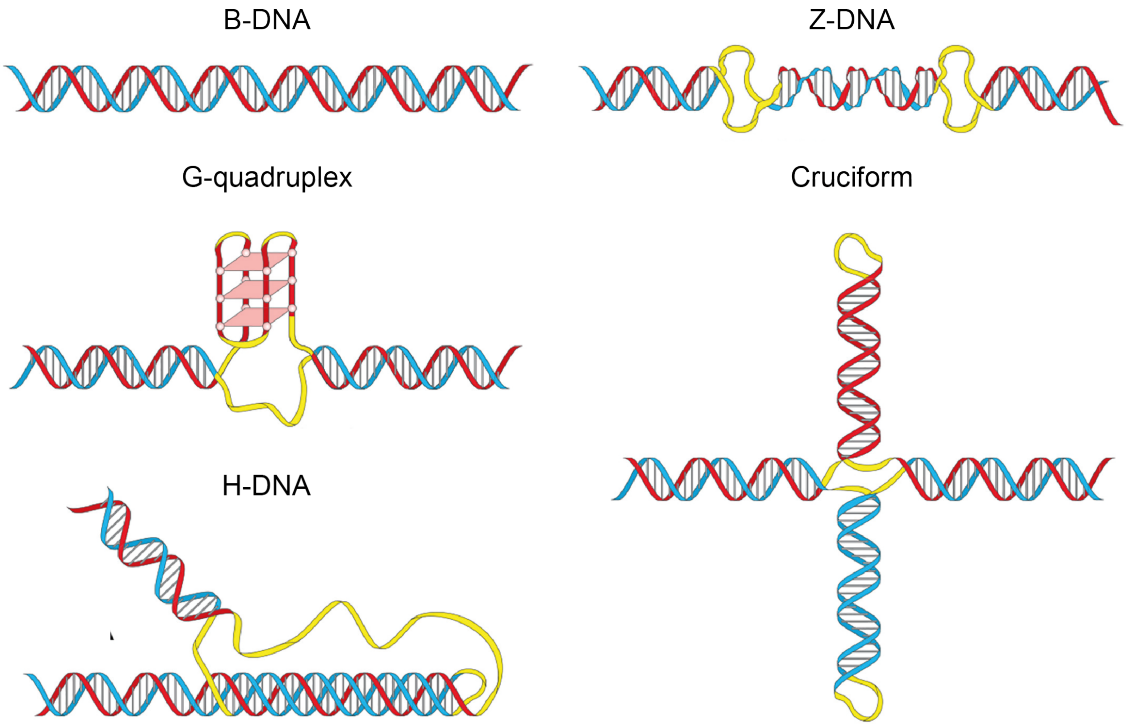

Figure 1.1: Canonical B-DNA and non-B-form DNA secondary structures.
(Top left) B-form DNA is the canonical right-handed double helix. (Top right) Z-DNA is a left-handed double helical structure that form in alternating purine-pyrimidine sequences. (Middle left) Four-stranded DNA structure formed from stacks of planar arrangements of four Hoogsteen-hydrogen bonded guanines. (Bottom-left) H-DNA is a three-stranded non-canonical DNA helix in which a single-stranded DNA binds to the major groove of DNA duplex. (Bottom-right) Cruciform DNA is a four-armed non-B-DNA structure, which has inverted repeated sequences repeat extruded to form stem loops. Figure adapted from Kouzine et al., 2017.

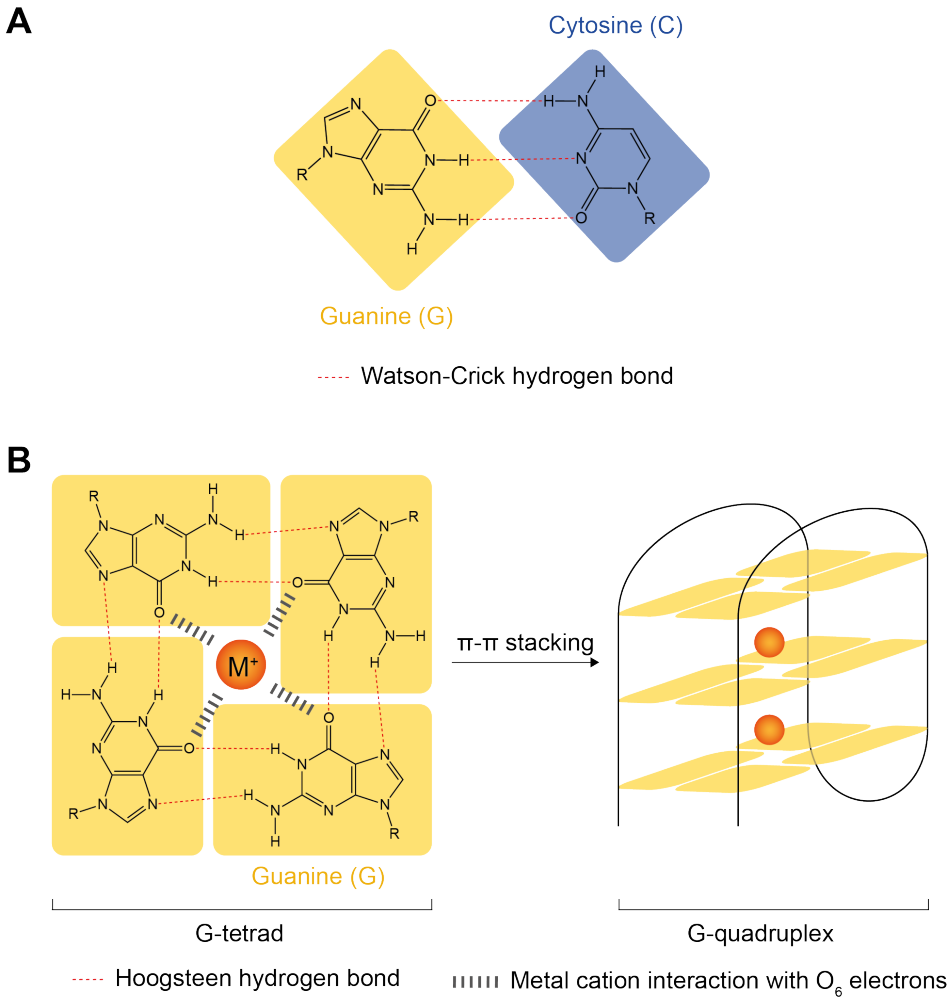

Figure 1.2: Structures of G-tetrads and G-quadruplexes.
(A) Watson-Crick hydrogen bonding between guanine and cytosine in canonical DNA duplex. **(B)** Schematic of a guanine (G)-tetrad and a G-quadruplex. Four guanines associate via Hoogsteen base-pairing to form a planar G-tetrad. A central monovalent cation (M^+) stabilises a G-tetrad by interacting with guanine O_6 electrons. A G-quadruplex is then stabilised by the stacking of multiple G-tetrads via **π-π** interaction. The stacks of G-tetrads are linked by loops of nucleotides (black line).

1.1.2 G4s are topologically polymorphic and heterogeneous

The G4 consensus sequence $G_{\geq 3}N_xG_{\geq 3}N_xG_{\geq 3}N_xG_{\geq 3}N_x$, which is derived from biophysical structural and computational analysis, has been proposed as a G4 motif (Eddy and Maizels, 2006; Huppert and Balasubramanian, 2005; Todd et al., 2005). Within the G4 motif, G-tracts (G≥3) are linked by nucleotide loops (N) of different lengths (X). Although G4 structures share similarities in primary sequence, they are a diverse family and highly heterogeneous folding into a variety of topologies.

One source of G4 heterogeneity is because of the strand(s) harbouring the G-tracts. G-quadruplexes can be assembled from one or several DNA strands. Genes containing a continuous G4 consensus sequence with three or more G-tracts can form intramolecular G4 structures, whereas intermolecular G4s can arise from two or four different G-rich strands, known as bimolecular G4s and tetramolecular G4s respectively (Burge et al., 2006; Murat and Balasubramanian, 2014) (Figure 1.3). Moreover, DNA-RNA hybrid intermolecular G4s have also been reported, and have been suggested to constitute an alternative stable structure (Wanrooij et al., 2012; Xiao et al., 2014; Zhang et al., 2020; Zheng et al., 2013).

G4 structural diversity also arises from the variety of strand polarity and strand orientation (Burge et al., 2006; Phan, 2010). The four backbone strands that form an interconnected G-tetrad can be assembled in parallel, antiparallel or even a mixture of parallel and antiparallel orientations (Figure 1.3). This variety of possible topologies also contributes to variability in intervening loop topologies, thought to be another source of G4 structural polymorphism. While early studies suggested that the relatively simple $G_{3\text{-}5}N_{1\text{-}7}G_{3\text{-}5}N_{1\text{-}7}G_{3\text{-}5}N_{1\text{-}7}G_{3\text{-}5}N_{1\text{-}7}$ consensus sequence forms the most stable G4s and that parallel conformation is favoured when one or more loops is short (Bugaut and Balasubramanian, 2008; Hazel et al., 2004; Piazza et al., 2015), more recent biophysical studies have shown that intervening loops with up to 21 nucleotides permit G4 folding (Bourdoncle et al., 2006; Guédin et al., 2010). Furthermore, a nuclear magnetic resonance (NMR) solution structure revealed the occurrence of a single-residue bulge, which is caused by an interruption of a single nucleotide between the guanines within the G-tracts, in G4 structures and extended the list of possible G4 topologies (Mukundan and Phan, 2013).

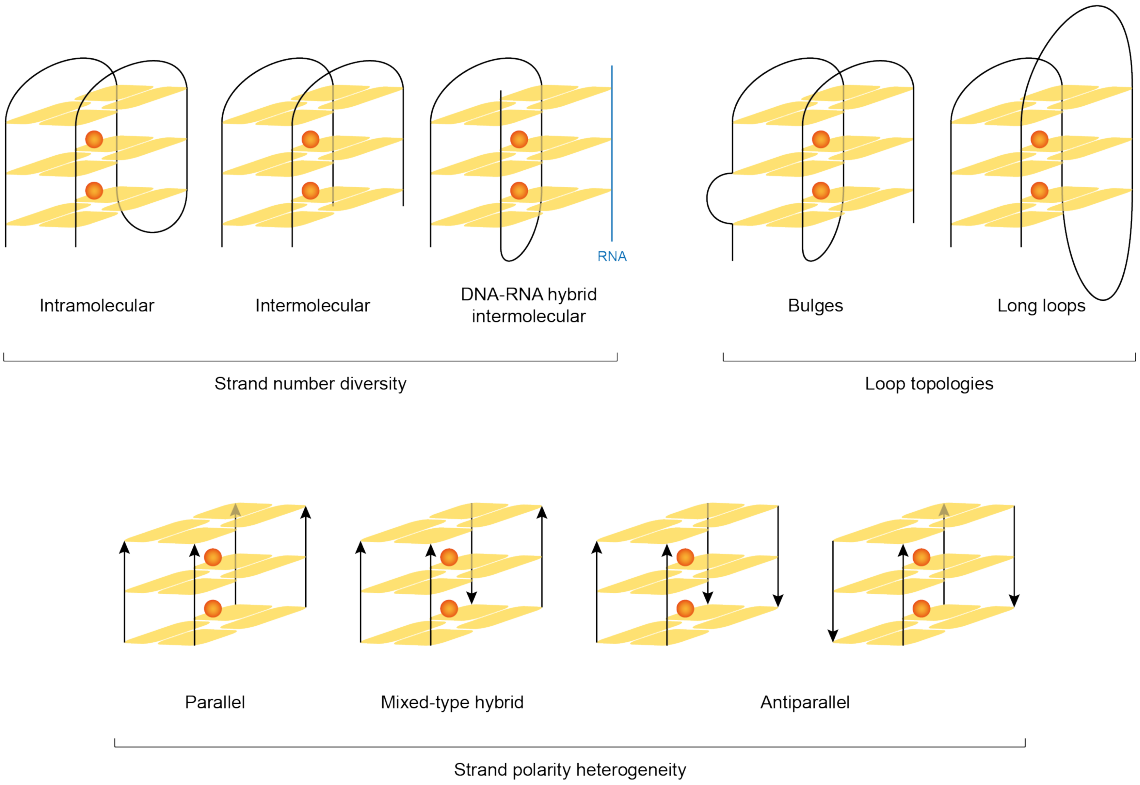

Figure 1.3: Topologies of G-quadruplexes.
Schematic representation of a series of G4 topologies differing in strand numbers and strand polarity. G4s can be assembled from one (intramolecular) or more (intermolecular) DNA strands. G-rich DNA can also interact with RNA strand to form hybrid intermolecular G4s. Depends on the orientation of backbone strands, G4s can be parallel, antiparallel or hybrid (i.e. a mixture of parallel and antiparallel). Length of the nucleotide loops can vary and bulges can exist if there is extra nucleotide(s) interrupt between guanines within the tract. These all contribute to G4 topological polymorphism.

1.2 Characterising G4 structures *in vitro*

1.2.1 Biophysical approaches to study G4s

The folding of G4 motifs in DNA or RNA oligonucleotides into G4 structures can be demonstrated by biophysical methods, and many of these experimental techniques have been used to investigate the structural properties of G4s. For example, circular dichroism (CD) spectroscopy is a common method used to determine and discriminate between G4 topologies (Baumann, 2010; del Villar-Guerra et al., 2018). In general, G4s with parallel conformations show a peak at 260 nm and a trough at 240 nm, while antiparallel G4s have a peak at 295 nm and a trough at 260 nm. Ultraviolet (UV) spectroscopy is another approach to characterise G4 structures, which is based on the fact that the UV light absorbance for different molecules varies. A UV signal at 295 nm marks a G4 structure, while a decrease in UV absorbance at the same wavelength suggests the melting of G4 structures (Mergny et al., 1998). A UV melting experiment, which employs UV spectroscopy in conjunction with thermal denaturation, can be used to measure G4 melting temperature and provide information about G4 thermal stability (Mergny and Lacroix, 2009; Mergny et al., 2005). The thermal stability of G4s can also be measured by a Förster resonance energy transfer (FRET) melting assay. In this assay, the 5' and 3' ends of the G4-forming oligonucleotide are respectively labelled with a donor and an acceptor fluorophore. When G4 folds (i.e. the donor and acceptor fluorophore are in proximity), the donor transfers energy to the acceptor and result in the fluorescent emission of the acceptor (Baumann, 2010; Ying et al., 2003). By measuring G4 melting temperatures, FRET can give information about G4 stability in different conditions (De Cian et al., 2007a).

X-ray crystallography can capture G4 conformations at atomic resolution (Campbell and Parkinson, 2007; Huppert, 2008). In addition to G4 conformation, X-ray crystallography also provides details of inter- and intramolecular interactions and ligand binding. This technology has been widely used in characterising ligand-bound quadruplex structures, as the X-ray diffraction data infers hydrogen bonding and electrostatic interactions between G4 and ligands (Campbell et al., 2012). To date, over 50 crystallised G4 structures, either ligand-bound or alone, have been characterised. However, X-ray crystallography only reveals structures in a solid state, which may not be the same as in solution. For polymorphic

structures, X-ray crystallography generally captures the form that is easy to be crystallised, rather than the form that is naturally favoured. An alternative method for high-resolution characterisation of G4 structures is NMR spectroscopy, which analyses G4s in solution and provides insights into their structure dynamics (Webba da Silva, 2007). Notably, an NMR spectroscopy study showed that in a G4 structure formed by a G-rich sequence in human *KIT* promoter, an isolated G-base predicted to be in the loop also participates in the forming G-tetrad core (Phan et al., 2007). This unusual feature, which has been shown to be stable in the K^+ solution, extends our growing appreciation of possibilities for polymorphism amongst G4 topologies. However, the existence of such features in cells requires further investigations.

Besides the techniques mentioned above, other approaches have also been adopted to identify G4 structures. For example, through the use of optical tweezers in conjunction with CD, Yu et al. showed the coexistence of parallel and antiparallel G4s in the promoter of insulin linked polymorphism region (ILPR), and these G4s were mechanically stable (Yu et al., 2009). G4 structures have also been studied by electrospray mass spectrometry (Ghosh et al., 2021), and the interactions between G4 and G4-specific drug have also been reported in this study. Overall, although biophysical techniques are widely used to study both DNA and RNA G4 structures, they are limited to the study of short oligonucleotides. Therefore, these methods can only reveal structures *in vitro* and do not account for the effect of flanking regions around G4 structures.

1.2.2 Biochemical methods to study G4s

Chemical and enzymatic approaches have also been developed to characterise G4 structures. Dimethyl sulfate (DMS) footprinting, which is also used in Maxam-Gilbert sequencing, is one biochemical method to characterise G4s (Sun and Hurley, 2010; Sun et al., 2011). G4 folding in stabilising conditions such as K^+ solution protects the guanine N_7 from being methylated by DMS and to subsequent depurination then DNA backbone cleavage by hot piperidine. Thus, after performing denaturing PAGE (polyacrylamide gel electrophoresis) analysis, the absence of cleavage patterns at the critical Gs involved in tetrad formation is consistent with G4 formation (Figure 1.4A). DMS has been useful to probe G4 conformations. For example, Siddiqui-Jain and his colleagues demonstrated that the G-rich

strand in the P1 promoter of *MYC* can form two different intramolecular G4 structures (chair-form and basket-form) by using DMS footprinting (Siddiqui-Jain et al., 2002). Destabilisation of chair-form G4 by a single G to A mutation resulted in increase of *MYC* transcription, while destabilisation of basket-form G4 did not change the transcriptional activity of *MYC*. This suggests that only the chair G4 is biologically relevant. Moreover, DMS footprinting followed by ligation-mediated polymerase chain reaction (LM-PCR) enables the detection of G4 structures in cells. In this method, genomic DNA is purified and cleaved by hot piperidine following DMS treatment in cells. LM-PCR is then performed to detect DMS footprints in regions of interest. By using this technique, Sun et al. demonstrated the existence of G4 structures in the *VEGF* (vascular endothelial growth factor) promoter in cells (Sun et al., 2011).

A DNA polymerase stop assay is also commonly used for G4 detection. The formation of G4s in the DNA template provides an obstacle to DNA synthesis by DNA polymerases (Han et al., 1999; Woodford et al., 1994). As a consequence, G4 structures can be detected *in vitro* by comparing the polymerase pause sites in G4-stabilising conditions (e.g. in the presence of K^+) with the original synthesised sequences in conditions that do not stabilise G4s (e.g. in the presence of Li^+) (Dexheimer et al., 2006; Guo et al., 2007; Qin et al., 2007) (Figure 1.4B). More recently, coupled with high-throughput sequencing and computational analyses, this assay allows genome-wide identification of sites that can potentially form G4s (G4-seq; see section 1.2.4) (Chambers et al., 2015). One limitation of the polymerase-stalling assay however, is that it does not provide characterisation of G4 structural conformations. Furthermore, this approach is generally employed for *in vitro* studies of G4 structures.

Similar approaches can also be applied to probe RNA G4s *in vitro* (Figure 1.4C-D). For instance, as reverse transcriptase pauses at RNA G4 structures, a reverse transcriptase stalling assay was developed to study both site-specific RNA G4s and genome-wide RNA G4s (rG4-seq) (Kwok and Balasubramanian, 2015; Kwok et al., 2016a). Furthermore, 2-methylnicotinic acid imidazolide (NAI) selectively acylates 2'-hydroxyl group of the unconstrained nucleotides and the 2'-OH acylation stalls the reverse transcriptase activity (Wilkinson et al., 2006). The newly developed SHALiPE approach, which couples selective NAI-mediated 2'-hydroxyl acylation with primer extension in the presence of Li^+, enables the detection of RNA G4s at single nucleotide resolution (Kwok et al., 2016b).

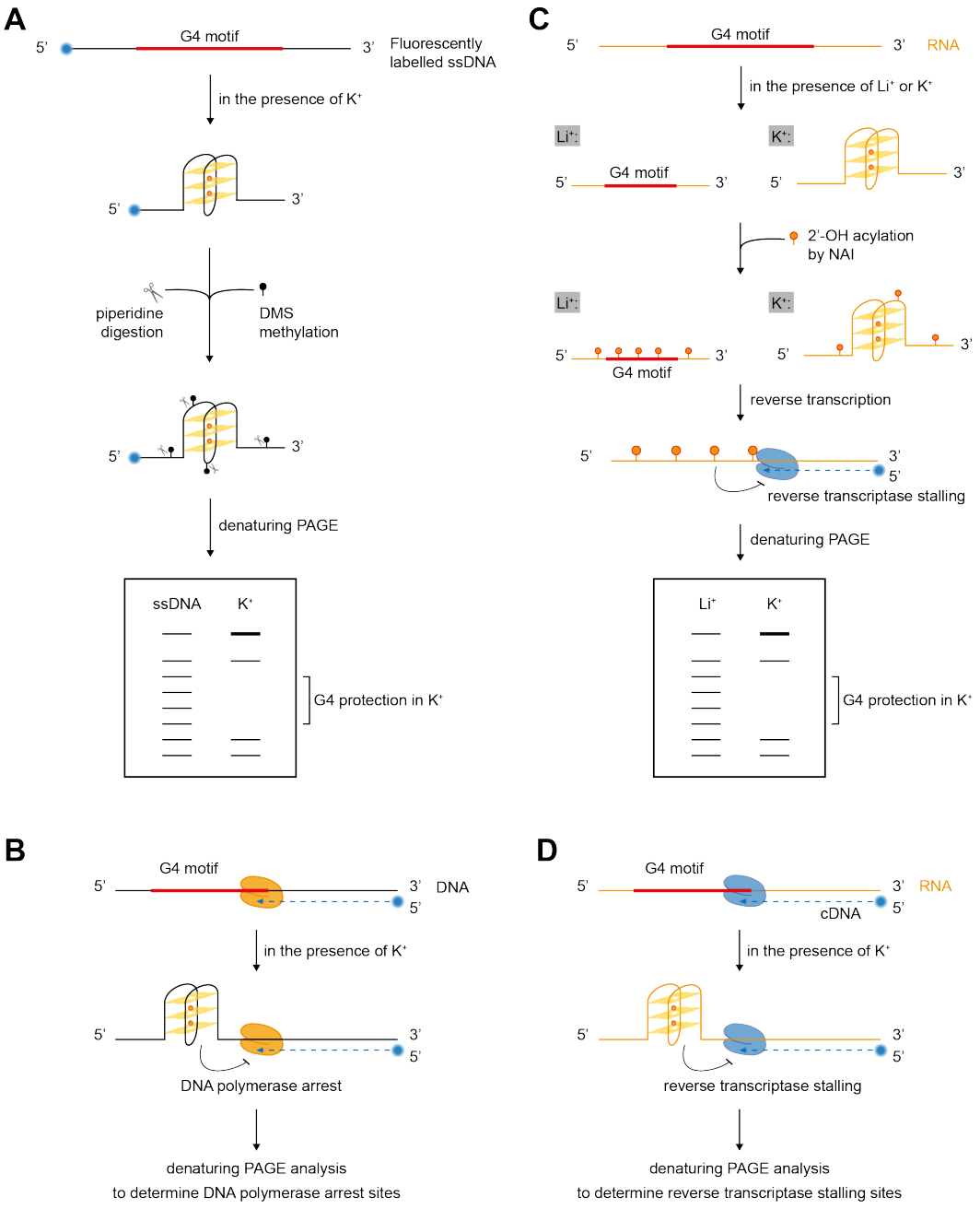

Figure 1.4: Biochemical techniques to detect DNA and RNA G-quadruplexes.
(A) DNA G4s are identified by dimethyl sulfate (DMS) footprinting, where DMS selectively methylates the N_7 of guanine residues, allowing its digestion by piperidine. The detection of G4 structures is subsequently analysed by denaturing PAGE. The absence of the bands suggests the protection of G4 formation. **(B)** Detection of DNA G4s by DNA polymerase stalling. DNA is synthesised in the presence of K^+ to detect DNA G4-dependent arrest sites. **(C)** RNA G4s are characterised by SHALiPE, which compares the differences in RNA 2'-OH acylation caused by 2-methylnicotinic acid imidazolide (NAI) and the reverse transcriptase stalling sites between G4-resolving conditions (e.g. Li^+) and G4-favouring conditions (e.g. K^+), to determine the formation of RNA G4s by denaturing PAGE. The absence of the bands suggests the protection of G4 formation. **(D)** Characterisation of RNA G4s by reverse transcription extension stalling. RNA is reverse transcribed in the presence of K^+ to identify RNA G4-dependent reverse transcriptase stalling.

1.2.3 Computational study of G-quadruplexes

Computational algorithms have been employed to predict the presence of potential G4-forming sequences in the human genome. One such algorithm is called Quadparser (Huppert and Balasubramanian, 2005). It predicts that more than 350,000 G4 motifs, denoted as putative quadruplex sequences (PQSs), in the human genome that are likely to adopt a G4 structure by searching for the consensus sequence $G_{3–5}N_{1–7}G_{3–5}N_{1–7}G_{3–5}N_{1–7}G_{3–5}$. Similar results were also obtained by other computational analyses (Eddy and Maizels, 2006; Kikin et al., 2006; Todd et al., 2005). Notably, these studies indicate that genomic PQSs do not occur randomly in the genome. Quadparser reveals that promoters are significantly enriched for PQSs, and more than 40% of promoters have one or more PQSs (Huppert and Balasubramanian, 2007). Promoter PQSs are overrepresented at genes involved in biological processes including cell signalling, development and cell differentiation (Eddy and Maizels, 2006; Huppert, 2008). Of particular interest, oncogenes are more likely to have a promoter PQS, with 67% harbouring a consensus G4 sequence. Some of the proto-oncogene promoters containing PQSs, include *MYC*, *BCL2*, *VEGF* and *KIT*, have been extensively characterised by biophysical (Agrawal et al., 2013; Mathad et al., 2011; Wei et al., 2015) and/or biochemical methods (Siddiqui-Jain et al., 2002; Sun et al., 2011).

Since early algorithms search relatively simple consensus sequences, such tools are unable to account for structural variants such as bulges, long loops and two-tetrad G4s and for the genomic contexts that influence G4 thermodynamic stability. To address some of these issues, more recent computational methods additionally incorporate analysis of the possibilities of higher-order assemblies and G-richness and G-skewness (Bedrat et al., 2016; Belmonte-Reche and Morales, 2020; Stegle et al., 2009). For example, G4Hunter takes G-richness and G-skewness of the sequence into account and provides a score of G4 propensity as an output (Bedrat et al., 2016). This model identifies 2- to 10-fold more sequences that are likely to be capable of forming G4s in human genome. Similarly, G4RNA screener provides a motif-independent method to identify potential RNA G4s *in vitro* (Garant et al., 2017). However, these algorithms have the incidence of false negatives (i.e. G4s that can form *in vitro* but escapes the computational prediction) and false positives (i.e. sequences that are predicted by algorithms but do not form G4s *in vitro*) (Guédin et al., 2009, 2010). Moreover, there remains a need for methods that consider chromatin context and protein-DNA interactions.

1.2.4 G4-sequencing in the human genome

Extending the detection of G4s in human genome by computational methods, a genome-wide DNA polymerase stop assay using high-throughput sequencing (G4-seq) provided the first map of G4s with potential to form a structure in the human genome (Chambers et al., 2015). By comparing DNA polymerase stalled sites under G4-stabilising conditions (in sequencing buffer containing K^+ or G4-stabilising ligand) with the reads from conditions that disfavour G4 formation (in Na^+ solution) (Figure 1.5), G4-seq identified over 716,000 distinct regions that can fold into a G4 structure *in vitro*, termed observed G4 sequences (OQSs). Of note, besides canonical G4s that comprise the consensus sequence $G_{3-5}N_{1-7}G_{3-5}N_{1-7}G_{3-5}N_{1-7}G_{3-5}N_{1-7}$, sequences that do not conform to the canonical G4 motif were also found to form G4s *in vitro*. Many of them correspond to non-canonical G4 structures including long loops (loop length longer than 7 bases), two-tetrads (G4s comprises only 2 G-tetrads) and bulges, which are not identified by previous computational predictions. Conversely, not all G4 sequences predicted by algorithms were identified by G4-seq, emphasising the need for direct G4 mapping in the genome. Subsequently, whole genome G4 mapping studies employing G4-seq have been extended to 12 different species from prokaryotes to eukaryotes, which include widely studied model organisms (Marsico et al., 2019). These reference G4 maps, which highlight the diversity in G4 sequence composition and location *in vitro*, have revealed that the high enrichment of G4-forming sequences in gene promoters is particular to mammals such as human and mouse.

Analogous to G4-seq, which maps G4s in DNA, RNA G4 structures were mapped using reverse transcriptase stalling on poly(A)-enriched RNAs (rG4-seq) (Kwok et al., 2016a). More than 3,000 RNA G4s were identified by rG4-seq throughout the human transcriptome. These RNA G4s were found enriched in untranslated regions (UTRs) and associated with microRNA target sites. Altogether, these techniques disclose G4 forming sites *in vitro*, providing important G4 reference maps. However, it is important to note that maps generated *in vitro* do not account for the effects of cellular context, such as chromatin and associated proteins, may have on G4 formation and stability. There still remains the need to generate G4 maps in cells.

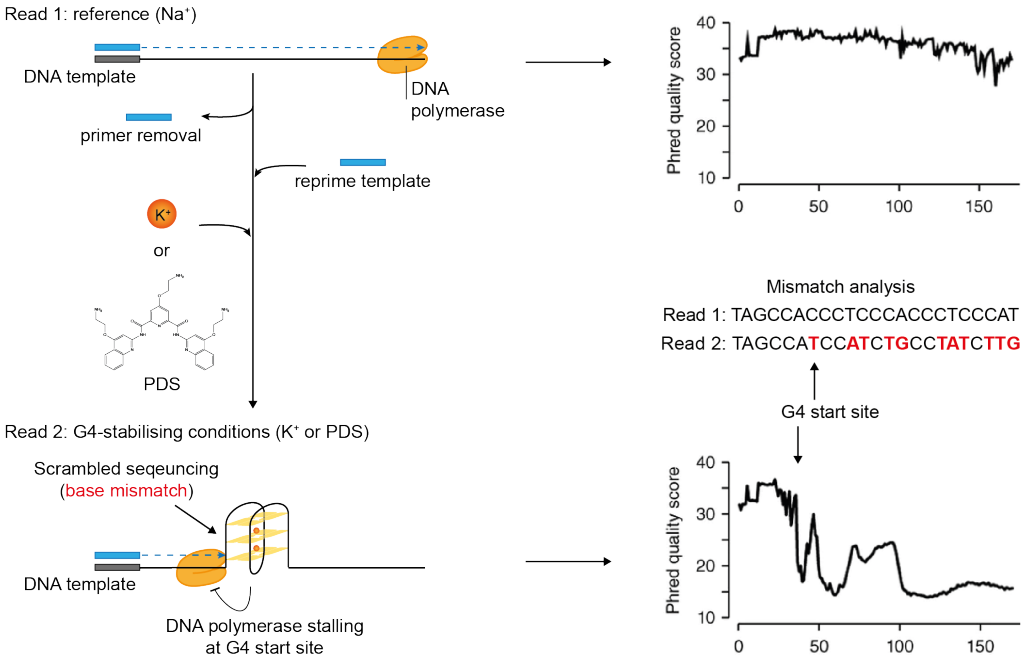

Figure 1.5: A schematic of G4-seq.
Sequencing is done twice in G4-seq. Read 1 provides sequencing reference in non-G4-forming conditions. Synthesised DNA is subsequently removed and the original DNA template is resequenced (read 2) under G4-stabilising conditions (in the presence of K^+ or PDS). Stalled polymerases result in a drop of sequencing quality and generate base mismatch (red) at the beginning of the G4 structure. Differences in base mismatches between read 1 and 2 are analysed to provide G4 map in the human genome. Figure is adapted from Chambers et al., 2015.

1.3 Tools and techniques to investigate endogenous G4 structures

1.3.1 G4-stabilising ligands

As G4s have been suggested to be enriched in functional regions of the human genome (Chambers et al., 2015; Huppert and Balasubramanian, 2007), many small molecules that recognise and stabilise G4s have been developed to explore the potential role of cellular G4s. G4 ligands generally have planar aromatic structures, which are thought to allow the stabilisation of G4s via the π-π stacking onto the terminal G-tetrad (Neidle, 2010) (Figure 1.6). For example, a natural molecule isolated from *Streptomyces anulatus* named telomestatin has been shown to facilitate the formation and stabilisation of intramolecular G4s, specifically telomeric G4s, and thus inhibit telomerase activity (Kim et al., 2002; Shin-ya et al., 2001). Pyridostatin (PDS) is a quadruplex stabiliser with both promoter and telomeric G4 affinity. Rodriguez et al. showed that PDS stabilised a G4 structure in proto-oncogene *SRC* (Rodriguez et al., 2012). When applied to human cancer cells, PDS induced replication- and transcription-dependent DNA damage at sites containing clusters of sequences that have the potential to form G4s, thus promoting growth arrest in human cancer cells (Rodriguez et al., 2012). The bisquinolinium compound PhenDC3, another widely used G4 ligand, inhibits replication fork progression *in vitro* and modulates gene expression in cells (De Cian et al., 2007b; Halder et al., 2012; Madireddy et al., 2016).

To date, around 1,000 G4 ligands have been reported in the G-quadruplex ligands database (G4LDB, http://www.g4ldb.org) (Li et al., 2013). The majority of them are reported as DNA G4-stabilising ligands since experimental characterisation was performed on DNA G4 oligonucleotides. It is important to consider however, that they may also bind RNA G4s. For instance, PDS also binds and stabilises RNA G4s both *in vitro* and in cells (Biffi et al., 2013a; Kwok et al., 2016a). Moreover, one PDS analogue, carboxyPDS (cPDS) (Figure 1.6), selectively interacts with RNA G4s over their DNA counterparts (Di Antonio et al., 2012).

G4-binding ligands have also been used to detect G4 structures in cells. G4 structures in human cells were visualised by treatment with derivatives of PDS and PhenDC3 followed by cell fixation and subsequent conjugation of fluorophores to the small molecules via click chemistry (Lefebvre et al., 2017; Rodriguez et al., 2012). G4 ligands that are intrinsically fluorescent, such as IMT and DAOTA-M2, have been developed for real-time G4 detection in

living cells since such ligands show different fluorescent emission or excitation wavelength and increased fluorescent lifetimes upon binding to G4s when compared to duplex or single-stranded DNA (ssDNA) (Shivalingam et al., 2015; Zhang et al., 2018). More recently, the existence of G4 structures in living cells was visualised using a fluorophore-tethered PDS analogue (SiR-pyPDS) for single-molecule imaging (Di Antonio et al., 2020). The study used SiR-pyPDS, which at a relatively low concentration (nM) that does not concomitantly induce G4-folding as seen by FRET analysis, to detect individual G4s and revealed natural structural fluctuation of G4s, suggesting that G4 folding is dynamic in living cells. Moreover, G4s in living cells have also been visualised using a fluorescent probe DAOTA-M2 in conjunction with fluorescence lifetime imaging microscopy (FLIM) (Summers et al., 2021). In this study, Summers et al. demonstrated that loss of FancJ and RTEL1 in mammalian cells led to the increase of the DAOTA-M2 lifetime, which suggested the increase of G4s. This indicates the role of FancJ and RTEL1 in resolving G4 structures in cells. Although G4 ligands allow real-time live cell imaging, the genomic location of the binding event in such studies is missing.

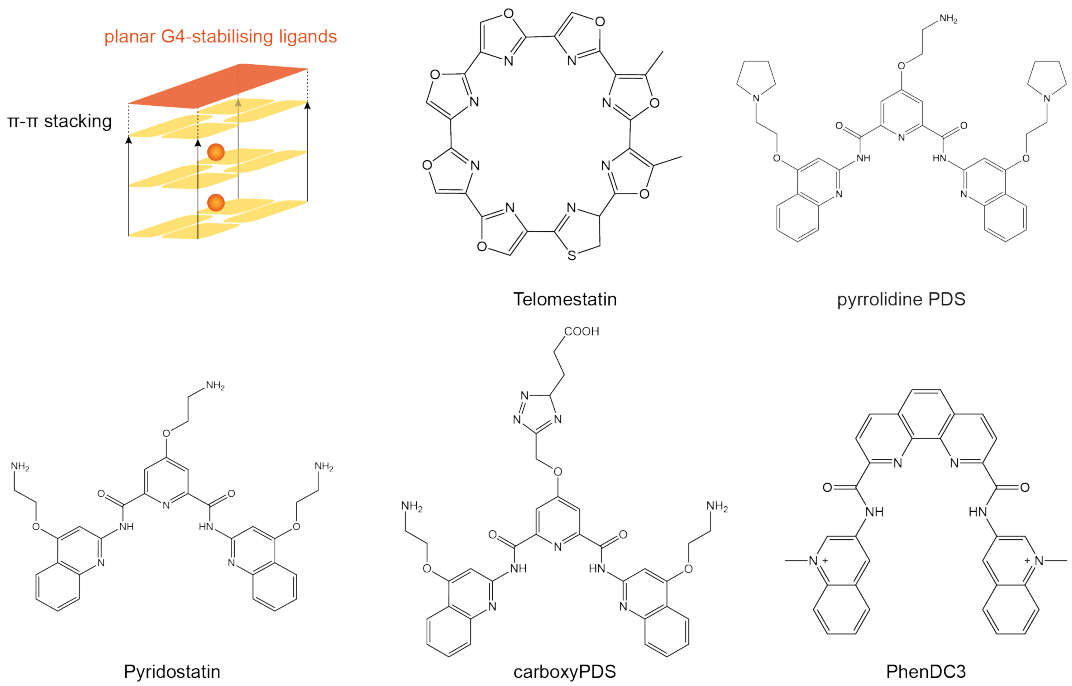

Figure 1.6: Examples of G4-stabilising ligands.
Planar aromatic small molecules interact with the terminal G-tetrad via **π-π** stacking to stabilise G4 structures (top left). The molecular structures for some examples of G4-stabilising ligands are provided.

1.3.2 Development of G4-specific antibodies

An alternative set of tools that have been developed to probe G4 structures in cells are antibodies that specifically bind to G4 structures. The single-chain variable fragment antibody (scFv) Sty49, which was derived via ribosome display and raised against telomeric G4s from *Stylonichia lemnae*, provided the first direct evidence that G4s exist *in vivo* (Schaffitzel et al., 2001). *Stylonychia* possesses two morphologically and functionally different nuclei, the micronucleus and the transcriptionally active macronucleus. The Sty49 staining in a replicating macronucleus suggested that G-quadruplexes are resolved during DNA replication, as the Sty49 signal was not observed in the replication band, the region where replication and telomere elongation take place.

More recently developed antibodies enable the detection of G4s in human cells. For instance, immunofluorescence (IF) staining in fixed human cells using the scFv antibody BG4, which has high affinity for several G4 conformations *in vitro*, revealed G4 formation in telomeric and non-telomeric DNA (Biffi et al., 2013b) (Figure 1.7). RNA G4s were also visualised in the cytoplasm of fixed human cells via BG4 IF staining (Biffi et al., 2013a). Subsequently, using the BG4 antibody, G4 structures were quantified in human tissues by immunohistochemistry, revealing that G4 formation is enriched in human stomach and liver cancer tissues compared to non-neoplastic tissues (Biffi et al., 2014).

The use of other G4-specific antibodies has also corroborated the finding that G4 structures exist in human cells. G4s were observed by immunostaining with the monoclonal antibody 1H6 and the scFv D1 (Henderson et al., 2014; Liu et al., 2016), and about 300 G4 forming regions were enriched following immunoprecipitation of purified and fragmented genomic DNA from human breast cancer cells using hf2 scFv (Lam et al., 2013). Nevertheless, the specificity of these antibodies may not be absolute. For example, the monoclonal antibody 1H6 has been shown to strongly bind to G4 structures with a $(T_4G_4)_2$ sequence but not bind to G4s without thymidines in the sequence, suggesting a sequence-dependent binding event for 1H6 (Kazemier et al., 2017; Liu et al., 2016). Moreover, sensitivity for detecting a single G4 structure over G4 clusters when applied to immunostaining remains unknown.

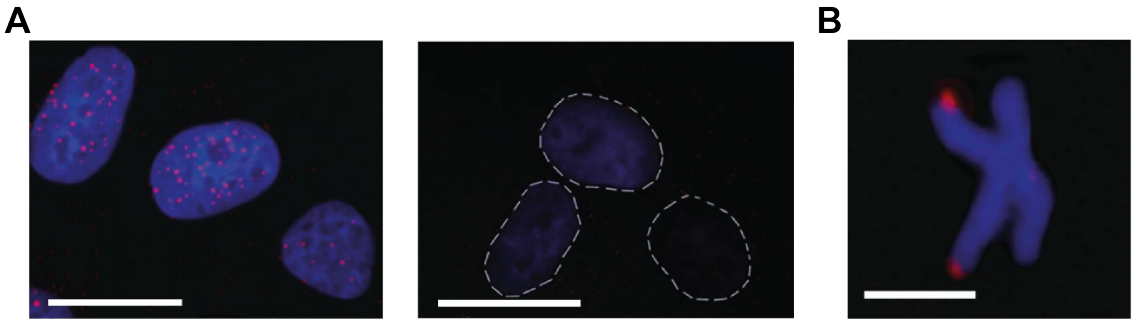

Figure 1.7: Visualisation of G4s in human cells using scFv antibody BG4.
(A) Immunofluorescence staining showing BG4 foci (red) in cell nuclei (blue) (left) and the absence of BG4 foci after DNase I treatment (right). Dotted lines show the boundary of nuclei. **(B)** Immunofluorescence staining showing BG4 foci (red) at telomeres of metaphase chromosomes. Scale bar, 20 μm. This figure is adapted from Biffi et al., 2013b.

1.3.3 High-throughput sequence mapping of endogenous G4s

To further investigate the formation of G4s at specific genomic loci within an endogenous chromatin context, efforts have been made to adapt high-throughput next generation sequencing tools to probe native G4 structures in cells. DNA G4s in chromatin from human cells have been mapped using BG4 antibody-based chromatin immunoprecipitation followed by next generation sequencing (G4 ChIP-seq) (Hänsel-Hertsch et al., 2016) (Figure 1.8). G4 ChIP-seq has identified 9,365 G4 structures in immortalised human epidermal keratinocyte HaCaT cells and 1,195 G4s in their normal counterparts, NHEK cells. These sites predominantly overlap with human G4 reference map identified by G4-seq as polymerase stall sites in purified human genomic DNA *in vitro* (Chambers et al., 2015), cross-validating the two methods. The number of G4 sites detected by G4 ChIP-seq is substantially (~70-fold) lower than the number observed by G4-seq, suggesting that G4 formation in cells can be governed by potential factors, such as local chromatin context and protein binding. By contrast, G4 ChIP-seq peaks are more numerous than the BG4 foci detected by IF staining (Biffi et al., 2013b), suggesting that the ChIP-seq method has higher sensitivity and resolution.

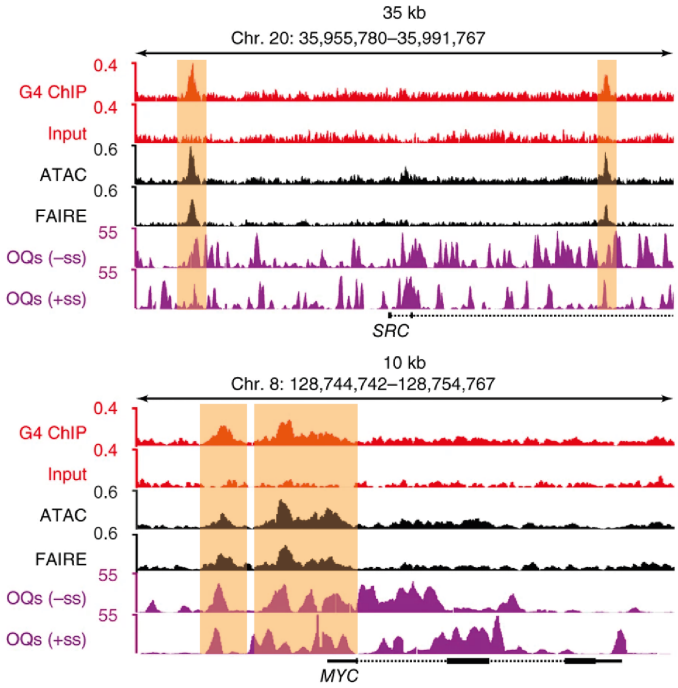

Figure 1.8: Endogenous G4 structures in human chromatin are detected by G4 ChIP-seq.
Example genome browser view for *SRC* and *MYC*. Tracks shows G4 ChIP-seq (top red) and input (bottom red) in HaCaT cells; accessible chromatin sites determined by ATAC-seq and FAIRE-seq (black); and reference G4 sites that can form G4 structures on the forward (+ss) and reverse (-ss) strand *in vitro* observed by G4-seq (OQs; purple). G4 ChIP-seq peaks which are located in nucleosome-depleted regions are masked in orange. Genomic coordinates are indicated above. The figure is adapted from Hänsel-Hertsch et al., 2016.

As well as BG4, G4 ChIP-seq has also been performed using the parallel G4-specifc antibody D1, which is directly expressed in human cervical carcinoma cells (Liu et al., 2016). Over 8,000 peaks were identified by this method, with a strong overlap with computationally predicted G4s. Notably, about 15% of G4s identifed by ChIP-seq using D1 are unique; they are not found by G4 ChIP-seq using BG4. Differences between these datasets might be explained by differences in protein expression methods, antibody specificities and the cell type-dependency of G4 landscape.

Another approach by which the locations of G4 structures have been inferred is the use of antibodies against proteins that are known to bind G4s. For example, yeast genomic regions bound by DNA helicase Pif1 and telomere-binding factor Rif1 strongly overlapped with computational predicted G4s (Kanoh et al., 2015; Paeschke et al., 2011). Furthermore,

human helicases such as ATRX, XPB and XPD have been found to bind G4-folded oligonucleotides *in vitro*, and have been mapped to computationally predicted G4 motifs in chromatin from human cells using ChIP-seq (Gray et al., 2014; Law et al., 2010). However, these approaches do not directly show the existence of a G4-folded structure. More recently, Zheng et al. developed an artificial G4 probe (G4P) by engineering a segment from the G4-binding domain of DHX36, an DNA and RNA helicase that binds and unwinds G4s (Chen et al., 2018b; Heddi et al., 2015; Zheng et al., 2020). G4P shows high affinity and specificity for G4s *in vitro*. By expressing G4P in human living cells followed by ChIP-seq, over 120,000 G4 sites were detected in cells. Why more sites were identified by ChIP-seq with G4P than with BG4 is not clear, but it is possible that the small size of G4P plays a role; being much smaller than scFv antibodies (ca. 7 kDa for G4P and ca. 30 kDa for scFv antibodies), it may have better accessibility. It is also possible that the use of G4P to probe living cells (G4 ChIP-seq using BG4, by contrast, is performed on fixed chromatin) may be a source of difference. However, this work still needs further cross-validation.

Besides G4 mapping using antibodies or proteins, chemical mapping also allows G4 detection in cells. By using potassium permanganate-dependent single-strand nuclease (permanganate/S1 Nuclease) footprinting combined with high-throughput sequencing, ssDNA regions of the genome in mouse B cell chromatin were identified (Kouzine et al., 2017). By combining mapped ssDNA patterns with computational prediction of non-B-DNA sequence motifs, the authors inferred sites of G4 formation (~20,000 G4s) in chromatin. These G4 structures were found enriched at promoters. Furthermore, the data suggested that the formation of G4s influence the positioning and occupancy of the flanking nucleosomes.

Although several RNA immunoprecipitation-based techniques have mapped binding sites of RNA G4 binding proteins in the human transcriptome, and have shown that such binding sites comprise G4 motifs, these techniques do not provide proof of the existence of an RNA G4-folded structure (Herdy et al., 2018; Murat et al., 2018; Sauer et al., 2019). High-throughput chemical mapping of RNA G4s by DMS and SHAPE has challenged that by showing that RNA G4s are globally unfolded in eukaryotic cells (Guo and Bartel, 2016). This is opposed to the results demonstrated by BG4 IF staining that RNA G4 foci were observed

in cell cytoplasm (Biffi et al., 2013a; Conlon et al., 2016). However, it is notable that DMS can irreversibly trap the unfolded G4 state (Di Antonio et al., 2020) and thus shifting the dynamic equilibrium of structural states towards unfolded states. Mapping endogenous RNA G4 structures still awaits the development of techniques that use G4 structure-specific probes. Preliminary results from our lab have suggested that endogenous RNA G4s can be mapped by RNA immunoprecipitation (RIP) using the BG4 antibody (Varshney D., unpublished data). As the focus of this thesis is DNA G-quadruplexes, techniques used to detect RNA G4s are not further discussed here.

1.4 Biological relevance of G4 structures

1.4.1 G4s in DNA biology

1.4.1.1 Transcription

The presence of a K^+-dependent G4 structure was initially detected at the promoter of the chicken β-globin gene *in vitro* (Woodford et al., 1994). Subsequently, many *in vitro* studies have observed the formation of G4s in synthetic, single-stranded oligonucleotides based on sequences found in the promoters of human genes, such as *MYC* and *KRAS* (Cogoi and Xodo, 2006; Simonsson et al., 1998). Computational predictions suggested that G4 structures are a common feature in the human genome and are enriched in gene promoters (Huppert and Balasubramanian, 2007), raising the hypothesis that G4 structures play a role in regulating transcription. Genome-wide G4 landscapes mapped by G4-seq in 12 different species confirmed that sequences with G4-forming potential are enriched at mammalian gene promoters. This enrichment is conserved in mouse and *Trypanosoma brucei*, indicating the conservation of possible similarities in the G4 biology. In a chromatin context, G4 structures mapped by G4 ChIP-seq were also found to be enriched in promoters of active genes (Hänsel-Hertsch et al., 2016). Remarkably, G4 ChIP-seq revealed that endogenous G4 structures mapped predominantly to nucleosome-depleted regions (Figure 1.8) and marked sites of active transcription. These findings further support the link between G4s and transcription. As understanding the mechanistic basis for this relationship was a central goal of this project, this topic is discussed in more detail below (see section 1.6).

1.4.1.2 Replication and genome instability

During DNA replication, transient opening of the DNA duplex may expose a single-strand region containing G4 motifs, thereby favouring the formation of G4 structures. Genome-wide bioinformatic analysis has suggested that the majority of human replication origins (~80%) are proximal to the sites containing predicted G4 consensus motifs, and that increased number motifs are associated with more frequently used origins (Cayrou et al, 2012; Besnard et al, 2012). These suggest that G4s may support DNA replication initiation. However, the relationship between G4s and replication origins in human cells are unclear, as replication origins are difficult to define in humans and show extensive variability (Leonard & Méchali, 2013).

Conversely, since folded G4 structures are stable under physiological conditions and have been shown to arrest DNA polymerases *in vitro* (Weitzmann et al., 1996; Woodford et al., 1994), the existence of G4 structures may also present in replication as a potential impediment. Many independent studies subsequently provided evidence that this is true *in vivo* and suggested that G4 structures may act as 'roadblocks' during DNA replication. Early evidence for this came from a loss-of-function study of a DNA helicase. In dog-1 (FANCJ) helicase deficient *Caenorhabditis elegans*, G-rich repetitive DNA elements with G4-forming potential are deleted (Kruisselbrink et al., 2008), suggesting that DNA polymerase is unable to process G4-associated regions. An independent study, which used the same genetic system, showed that a single unresolved G4 structure persisted on the leading strand of DNA through mitotic divisions and imposed a new block during each subsequent replication process. Failed replication of the G4-forming site caused a single-strand DNA gap, which further induced double-strand DNA breaks in subsequent divisions (Lemmens et al., 2015).

By arresting the progress of DNA polymerase during replication, the formation of G4 structures can trigger replication stress, which can in turn cause DNA damage and lead to genome instability. Works in model organisms provide evidence that G4s are determinants of genome instability. In addition to the studies of FANCJ-deficient *C. elegans* mentioned above, Piazza et al. showed that the human CEB1 minisatellite, which harbours G4 motifs, becomes unstable when introduced in Pif1-deficient yeast or in yeast cultured in the presence of G4 ligands (Piazza et al., 2010, 2017). Furthermore, mutational analysis of

CEB25 in yeast suggests that increased G4 thermostability, which is achieved by shortening the G4 loop length (from 9 nt to less than 4 nt), causes greater genomic instability *in vivo* (Piazza et al., 2015).

1.4.1.3 Epigenetics

Emerging evidence also indicates that G4 structures interact with DNA modifications and histone modifications. During replication stress following depletion of nucleotide pools in chicken DT40 cells, heritable epigenetic reprogramming was induced by small molecule mediated G4 stabilisation or loss of the REV1 protein, which is required for replication of G4-forming sequences (Guilbaud et al., 2017; Papadopoulou et al., 2015; Sarkies et al., 2010). This included the loss of H3K4me3 and subsequent cytosine methylation. In addition, a G4 structure at the hTERT promoter has been shown to recruit NME2 and REST-LSD1 to remove gene-activating H3K4me1 and H3K4me2, thus resulting in hTERT repression (Saha et al., 2017). The transcription of *CDKN1A* is regulated by a similar mechanism, in which a TRF2-G4 interaction is required for REST-LSD1 recruitment (Hussain et al., 2017). DNA methyltransferases (DNMTs), which methylate cytosine at C-5 and are predominantly located at CpG dinucleotides in mammalian cells, also showed preferential binding to G4-forming DNA over double-stranded DNA *in vitro* as defined by protein binding assays (Cree et al., 2016; Mao et al., 2018). Furthermore, the binding of DNMT1 to G4 structures inhibits it enzymatic activity, and the presence of G4 structures in cells is highly associated with hypomethylation at DNMT1 binding sites (Mao et al., 2018) (Figure 1.9). This implies that G4 structures sequester DNMT1 to prevent CpG methylation and therefore mould the DNA methylome.

In addition to cytosine methylation, 8-oxo-7,8-dihydroguanine (8-oxoG), an oxidised DNA base damage in the genome, is also associated with G4s. G bases in predicted G4-forming motifs have low redox potential and are susceptible to 8-oxoG formation (Fleming and Burrows, 2017). It has been indicated that 8-oxoG within telomeric G4 structures disrupts G4 structures by reducing the thermal stability of G4s, and as a result stimulates telomerase function and promotes telomere instability (Fouquerel et al., 2016, 2019). Emerging evidence has also suggested that the interplay between 8-oxoG and G4s can regulate transcription. As shown in a luciferase reporter assay, the removal of 8-oxoG from a G4

motif on the coding strand of *VEGF* promoter by OGG1-mediated base excision repair (BER) unmasked the G4 folding and resulted in upregulated transcription (Fleming et al., 2017). In contrast, the removal of 8-oxoG of the G4 motif on the template strand turned off the transcription. Furthermore, a recent study observed that the binding sites of OGG1 and APE1, proteins that are involved in 8-oxoG-induced BER, were predominant at G4 motifs (Roychoudhury et al., 2020). Roychoudhury et al. also found that G4 formation was promoted by the binding of APE1 and APE1 subsequently facilitated G4-mediated gene expression.

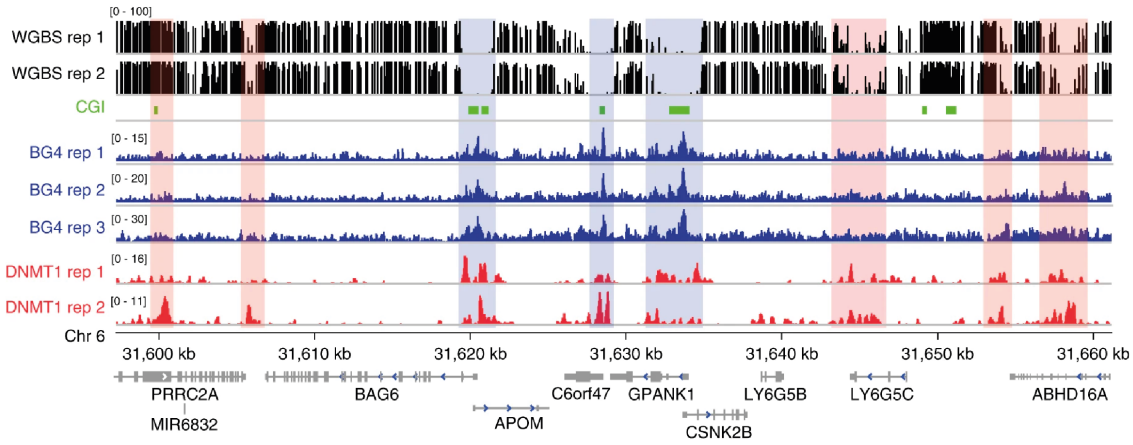

Figure 1.9: DNMT1 is recruited to the G4-forming sites associated with hypomethylation.
Example genome browser view shows tracks of whole genome bisulfite sequence (top, black), CpG islands (green), G4 ChIP-seq (blue) and DNMT1 ChIP-seq (red) at the hypomethylated region from Chromosome 6. Regions masked in blue shows that the presence of G4s is associated with DNMT1 binding and low methylation, while orange-masked regions are hypermethylated sites enriched with DNMT1 but do not possess a G4 structure. Figure is adapted from Mao et al., 2018.

1.4.2 G4s in RNA biology

To date, RNA G4s have been suggested to play multiple roles in RNA biology. For example, G4 structures have been found enriched in the beginning of 5'-untranslated regions (5'-UTRs), suggesting that they may have functions in translation (Chambers et al., 2015; Huppert et al., 2008). Many studies have reported the relationship between G4s and translation individually. For example, Kumari et al. has demonstrated that the folding of the RNA G4 in the 5'-UTR of the transcript of human *NRAS* proto-oncogene leads to translational inhibition *in vitro* (Kumari et al., 2007). Similarly, it has been shown that folding of the thermodynamically stable *BCL-2* RNA G4, which is well-characterised biophysically, inhibits translation both *in vitro* and in human cells (Shahid et al., 2010). Moreover, inhibiting the eukaryotic initiation factor 4A (eIF4A), which is the helicase that unwinds structured 5' UTRs and facilitates the recruitment of ribosomes to initiate translation, resulted in reduced translation efficiency of mRNAs with longer 5' UTRs enriched in two-tetrad G4s (Wolfe et al., 2014). This indicates that the formation of RNA G4s impedes recruitment or the scanning of ribosomes during translation initiation. The occurrence of RNA G4s in RNA coding regions is relatively less abundant compared to 5'-UTRs. Interestingly, a computational study has suggested that most stable G4s that are predicted to fold in coding sequences are largely lost through the use of synonymous codons during evolution (Mirihana Arachchilage et al., 2019). The interactions of RNA G4s with RNA-binding proteins also appear to have impacts on mRNA translation. Examples include the RNA-binding protein CNBP, which preferentially binds and unwinds G4-folding elements in mRNA coding sequences to support translation (Benhalevy et al., 2017).

In addition to translation, RNA G4s have also been suggested to play other roles in RNA biology, such as alternative splicing. The splicing factor hnRNPH has been shown to bind RNA G4s *in vitro*, and dysregulation of hnRNPH correlates with elevated RNA G4s (Conlon et al., 2016). Moreover, it has been suggested that RNA G4s promote alternative splicing through the binding of splicing factor hnRNPF, and exon inclusion is abrogated when the G4-forming capacity, but not the G-tract sequence, of an RNA is destroyed (Huang et al., 2017). By using a *NRAS* RNA G4-oligonucleotide as a bait for interacting proteins from cytoplasmic extracts, many RNA G4 interacting proteins were identified using liquid chromatography-tandem mass spectrometry (Herdy et al., 2018). These include RNA helicases DDX5 and

DDX17, which have been reported to have G4-relevant roles in splicing (Dardenne et al., 2014). More recently, accumulating evidence also suggests that RNA G4s may have a role in stress granule formation. For instance, RNA G4-binding proteins DHX36 and DDX3X are associated with stress granule formation (Chalupníková et al., 2008; Valentin-Vega et al., 2016). tRNA-derived stress-induced RNAs (tiRNAs) are tRNA fragments generated under cellular stress and are important in stress granule formation (Lyons et al., 2016; Tao et al., 2020). Lyons et al. observed that the formation of tetramolecular RNA G4s were assembled from four individual tiRNAs during cellular stress, and disruption of such RNA G4 structures abrogated the tiRNA-mediated stress granule formation in cells (Lyons et al., 2017).

1.4.3 G4s at telomeres

Mammalian telomeres at chromosome ends are composed of single-stranded tandem repeats of TTAGGG DNA sequences associated with protein shelterin complexes, which maintain genome integrity and protect telomeric DNA from DNA damage (Shay and Wright, 2019). These G-rich regions have been shown to form stable DNA G4 structures (Parkinson et al., 2002). It has been demonstrated that mammalian telomere DNA is transcribed into telomeric repeat-containing RNA (TERRA) (Azzalin et al., 2007; Schoeftner and Blasco, 2008), and TERRA is reported to form RNA G4s *in vitro* and in human cells (Martadinata and Phan, 2009; Xu et al., 2008, 2010). Moreover, telomeric DNA and RNA were found to assemble into a hybrid DNA-RNA G4 conformation (Bao and Xu, 2020; Xu et al., 2009). Therefore, various G4 structures may exist at telomeres and play a role in telomere biology.

To date, many studies have reported the relationship between G4s and telomere maintenance. For instance, Paeschke and colleagues have demonstrated the existence of antiparallel telomeric DNA G4s in the ciliate *Stylonychia lemnae* and shown that the maintenance of telomere G4 structures *in vivo* requires the ciliate telomere end-binding proteins (TEBPs), TEBPα and TEBPß (Paeschke et al., 2005). As telomeric G4s have been observed in human telomeres using BG4 IF and G4 ChIP-seq (Biffi et al., 2013b; Henderson et al., 2014; Liu et al., 2016), and telomere binding proteins Rif1 and TRF2 bind to G4s *in vitro* (Biffi et al., 2012; Kanoh et al., 2015), it has been suggested that G4s may be involved in telomere capping. Indeed, it has been found that in budding yeast that has mutated *cdc13-1*, the gene encodes the Cdc13 capping protein, the telomere capping is lost and the

growth inhibited (Smith et al., 2011). By the overexpression of DNA G4 binding proteins, treatment with G4-stabilising ligands or loss of the DNA G4-resolving helicase Sgs1, which were shown to stabilise telomeric DNA G4s *in vitro*, the growth of cdc13-1 mutants was rescued. This supports the idea that telomeric G4s can perform a rudimentary role in telomere capping.

As well as DNA G4s, RNA G4s also have roles in telomere biology. A G4 structure that folds in the long non-coding TERRA has been proposed to act as a scaffold in telomeres to maintain chromatin homeostasis (Takahama et al., 2013). Human telomere-binding proteins TRF2 and FUS can bind to both the TERRA RNA G4 structure and the telomere DNA G4 *in vitro* (Biffi et al., 2012; Takahama et al., 2013). This co-binding enables FUS and TRF2 to recruit histone methyltransferases which are critical for heterochromatin formation at telomeres (Takahama et al., 2013).

G4 structures have also been reported to have multiple roles in telomerase-mediated telomere extension. On the one hand, an *in vitro* study has shown that the folding of anti-parallel telomeric G4 structures inhibits *Oxytricha nova* telomerase and subsequent telomere elongation (Zahler et al., 1991), suggesting that telomeric G4 structure may negatively regulate telomere extension. On the other hand, parallel telomeric G4 structures can be partially resolved and extended by telomerase *in vitro* (Moye et al., 2015), and works in yeast and in human cells have shown that telomerases are recruited upon the stabilisation of parallel telomere G4s (Moye et al., 2015; Zhang et al., 2010a). These opposing functions of G4s in telomere extension may reflect evolutionary divergence in telomerase properties or complex regulatory roles of G4s in telomere biology.

1.4.4 G4-associated cancer and therapeutic opportunities

G4 structures appear to be enriched in functional regions of the genome, such as promoters and 5'-UTRs (Chambers et al., 2015; Hänsel-Hertsch et al., 2016; Huppert and Balasubramanian, 2007). Particularly striking is their enrichment in genes associated with cancer, and in regions that undergo somatic copy number alterations (CNAs) related to cancer development (Chambers et al., 2015; Georgakopoulos-Soares et al., 2017; Hänsel-Hertsch et al., 2016), raising the question of what the relationship is between G4s and

cancer. In support of this, several lines of evidence suggest that G4 structures preferentially exist in cancer states when compared to healthy states. For example, using the BG4 antibody to detect DNA G4 structures in different types of patient-derived tissues by immunohistochemistry, Biffi et al. found that the presence of DNA G4s was significantly higher in the nuclei of stomach and liver cancer tissues than the corresponding non-neoplastic tissues (Biffi et al., 2014) (Figure 1.10). Similar findings were reported in an independent study that compared immortalised cancer-like human epidermal keratinocyte HaCaT cells to their primary normal NHEK counterparts (Hänsel-Hertsch et al., 2016). Here, through the use of BG4 IF staining, researchers observed 4-fold more BG4 foci in HaCaT than NHEK cells. Consistent with this, G4 sites detected by G4 ChIP-seq in HaCaT cells were 10-fold higher than in NHEK cells. Furthermore, a recent study of the G4 landscape in chromatin from patient-derived aggressive breast cancer xenografts suggested that promoters of highly amplified genes, whose expression is increased, also harbour a greater number of G4s (Hänsel-Hertsch et al., 2020). The study also showed that differentially enriched G4s identified across 22 samples were more likely to be found at CNAs. These findings suggest that potentially important aspects of intratumour heterogeneity (i.e. differences of molecular and phenotypical profiles for cell populations within the same tumour) are reflected in the G4 landscape.

G4 structures are associated with a growing number of biological processes that are important in cancer development, including telomere biology, DNA replication, genome instability and cancer-related gene regulation. This has stimulated exploration of G4-targeting as a method of intervention for cancer therapy. Telomerase, which plays a role in the maintenance of telomere homeostasis, is overexpressed in the vast majority of human tumours (85-90%) (Kim et al., 1994; Shay and Wright, 2019). Since the formation of telomeric G4 structures inhibits telomerase activity (Zahler et al., 1991), attempts have been made to exploit inhibition of telomerase-mediated telomere extension by G4-stabilising ligands for cancer therapeutics (Sun et al., 1997). Subsequently, evidence emerged that some G4 ligands caused DNA damage at telomeres. One mechanism by which they appear to do so is via PARP1, a key regulator of the nonhomologous end joining (NHEJ) repair pathway. In both cellular and xenograft models, stabilisation of telomeric G4s by the G4 ligand RHPS4 specifically recruits PARP1 and activates the poly-ADP ribosylation, a post-

translational modification of proteins at telomeres, and RHPS4 treatment in PARP1 knockdown cancer cells inhibits cell growth (Salvati et al., 2010). By contrast, the displacement of the telomeric shelterin protein hPOT1 using the G4 ligand BRACO-19 also caused telomere DNA damage and subsequently cell death, suggesting potential applications for chemotherapeutics (Neidle, 2010). DNA damage can also be induced throughout the genome using G4 ligands. For example, the G4-interacting drug PDS induced DNA damage in human cancer cells and promoted cell death (Rodriguez et al., 2012).

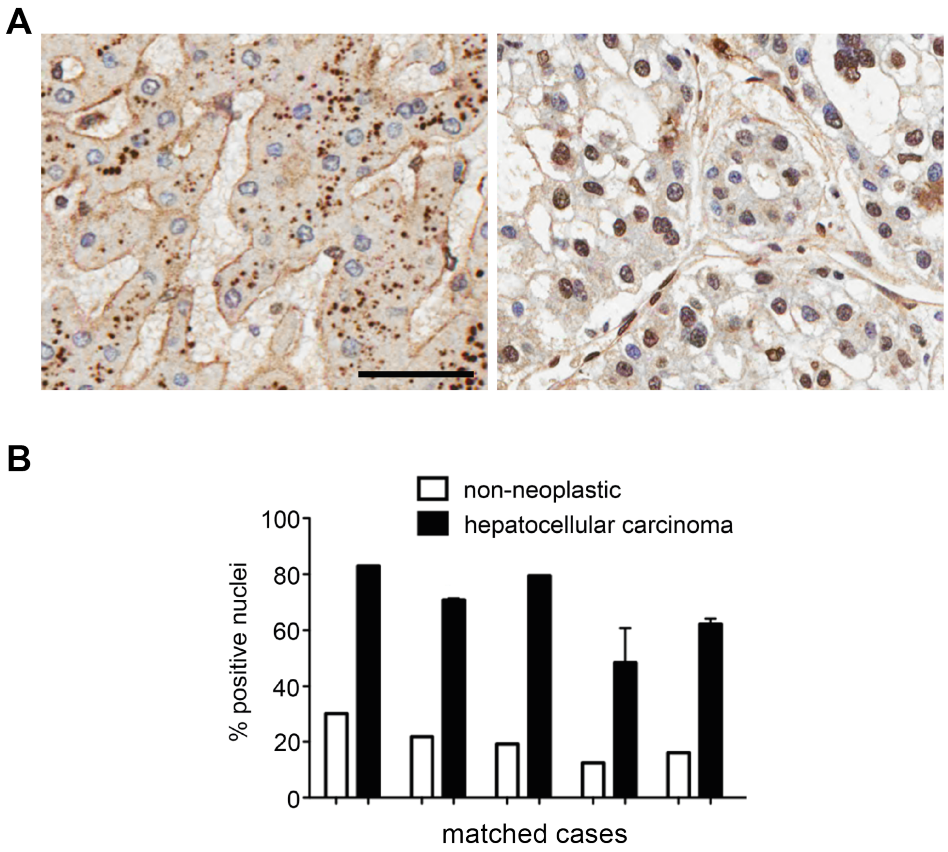

Figure 1.10: Human liver cancer tissue shows more DNA G4s than matched non-neoplastic tissue.
(A) Examples for liver cancer tissues (left) and non-neoplastic tissues (right) stained by immunohistochemistry using G4-specific antibody BG4. The nuclei of non-neoplastic liver tissue are mostly BG4-negative, with haematoxylin counterstaining (blue) evident, while neoplastic tissue shows the extensive presence of BG4-positive nuclei (brown). Scale bar, 50 μm. **(B)** Quantification of BG4 positive nuclei in tissues from 9 individual patients. Error bars represent the standard error of the mean (s.e.m.). This suggests that there are more DNA G4s in liver cancer compared to corresponding non-neoplastic tissue. Figure is adapted from Biffi et al., 2014.

Efforts have also been made to manipulate the transcription of G4-containing cancer genes by targeting their promoter G4s. Apart from PDS, which has been shown to inhibit the expression of numerous oncogenes (Rodriguez et al., 2012), multiple G4 ligands have shown promise for this purpose. One example is CM03. CM03-treated human pancreatic ductal adenocarcinoma (PDAC) cells showed a global reduction in the expression of many G4-containing genes, and the tumour growth of PDAC xenografts and KPC mouse models was arrested by CM03 without causing general toxicity (Marchetti et al., 2018). Another approach has been made to exploit the synthetic lethality (increased cell death in combination, but not alone) of G4 ligands and specific genetic changes found in cancer cells. For example, PDS treatment was shown to cause cell lethality in BRCA2-deficient cells, a gene involved in homologous recombination (HR) repair pathway (McLuckie et al., 2013). The study also demonstrated the synthetic lethality of cells treated with both PDS and NU7441, an inhibitor of the DNA-PK kinase crucial for NHEJ repair pathway. Furthermore, exposure of BRCA1/2 deficient tumours to CX-5461 causes cell lethality (Xu et al., 2017). CX-5461 was previously thought to be an RNA Pol I inhibitor, but was shown to be a G4 stabiliser and to induce ssDNA breaks, which require BRCA for homologous-directed DNA repair. More recently, a genome-wide shRNA screen in the presence of G4 stabilising ligands has revealed many novel genes and pathways that are required for cell survival specifically in the presence of G4 ligands (Zyner et al., 2019). For example, deficiency of WEE1 cell cycle kinase or USP1 deubiquitinase in the presence of G4-stabilising ligand promote cell killing. Moreover, helicase DDX42, which was reported as a newly discovered G4-binding protein in this study, was designated as a 'key' gene whose deficiency imparts significant G4 ligand sensitivity for a range of cell types and/or with a range of G4 ligands. In the future, overlapping this resource with genetic deficiencies found in cancer cells may point to as yet unexplored vulnerabilities and to new therapeutic possibilities.

1.5 Transcriptional regulation in eukaryotic cells

The work in this thesis aims to explore the interaction of G4s with transcription and chromatin state. To provide context to models that I propose in result chapters, I will briefly introduce some background to key principles of transcription machinery in this section.

1.5.1 Chromatin accessibility and active gene transcription

A key determinant of transcriptional output is the accessibility of regulatory elements. Chromatin accessibility is determined by the occupancy and topology of nucleosomes and chromatin-binding factors, non-histone macromolecules that bind to DNA directly or indirectly, that occlude accessibility of DNA (Klemm et al., 2019). The nucleosome is the fundamental element of chromatin, composed of an octamer of histone proteins wrapped with DNA (~147bp). Nucleosome occupancy is dynamic across the genome (Poirier et al., 2008), creating regions that range in accessibility from closed chromatin to permissive chromatin and highly dynamic open chromatin. Across the genome, the accessibility of chromatinised DNA is regulated by external stimuli and varies between cell types. Transcription factors (TFs) play critical roles in this regulatory process, through sequence-specific DNA-binding (Klemm et al., 2019). TFs gain access to histone-bound DNA during nucleosome turnover and displace nucleosomes to generate accessible regions for subsequent transcriptional regulators binding (Bao et al., 2015; Svaren et al., 1994; Workman and Kingston, 1992). Important exceptions to this are pioneer factors, TFs that are able to bind to nucleosomal DNA directly during the establishment of chromatin state and actively open local chromatin for active transcription (Zaret and Carroll, 2011). In addition to TFs, chromatin-remodeling enzymes play a central role in regulating chromatin state. For example, the SWI/SNF chromatin remodeler family disrupts the interaction between genomic DNA and nucleosomes by sliding and evicting nucleosomes, although they are not involved in chromatin assembly (Längst and Manelyte, 2015). By contrast, ISWI complexes contain a nucleosome recognition module that binds to DNA, and possess catalytic activity that regulates nucleosome spacing to either promote chromatin assembly or inhibit chromatin compaction (Gkikopoulos et al., 2011; Yang et al., 2006).

Chromatin states correlate with specific post-translational histone modifications (Bannister and Kouzarides, 2011). One important example is histone acetylation, which is critical for modulating chromatin structure. The acetylation of histones is regulated by two groups of histone modifying enzymes, histone acetyltransferases (HATs) and histone deacetylases (HDACs) (Bannister and Kouzarides, 2011; Wang et al., 2009) and results in decompaction of chromatin structure (Li and Reinberg, 2011; Robinson et al., 2008). The acetylation of lysine side chains in histones neutralises the lysine positive charge and destabilises the

interactions between DNA and histones, which is important for the preclusion of histone binding (Bannister and Kouzarides, 2011). For example, accumulating evidence has indicated that HAT1 acetylates histone H4 in cells, and depletion of HAT1 results in the decrease of acetylation in H4 (Barman et al., 2006; Parthun, 2007). Acetylation has also been found present within the histone core. GCN5 is a HAT protein that acetylates lysine 56 in histone H3 (H3K56), whose side chain points to DNA major groove (Parthun, 2007; Tjeertes et al., 2009). This suggests that acetylation of H3K56 also affects histone-DNA interactions. Furthermore, acetylation was also found to be associated with rapid nucleosome turnover (Wang et al., 2009), which in turn increases chromatin accessibility. Histone methylation is another important example of post-translational histone modification. Unlike histone acetylation, histone methylation does not have impact on the charge of histone proteins (Bannister and Kouzarides, 2011). Histone methylation occurs on the side chains of lysines and arginines. Amongst histone methylations, H3K4, H3K9, H3K27 and H3K36 are the most common and well-studied sites. These lysine methylations mark gene repression except H3K4me3 which marks accessible chromatin and active transcription (Barski et al., 2007; Hyun et al., 2017; Saksouk et al., 2015). Moreover, methylation of H3K9 marks heterochromatin regions via recruiting chromatin modifying proteins, such as the heterochromatin protein (HP) family, and is mediated by histone methyl transferases (Batie et al., 2018; Hyun et al., 2017; Ninova et al., 2019).

Increasing evidence indicates that the interaction between histone modifications and chromatin remodelers is crucial in regulating chromatin (Swygert and Peterson, 2014). For example, human BRG1, a SWI/SNF subunit, recognises the acetylation of H4K8 and H4K12 and activates interferon-ß gene transcription (Agalioti et al., 2002). In some cases, chromatin remodelers directly modify histones. The NURD remodeler complex has two HDAC subunits, which enable histone deacetylation to repress gene transcription (Xue et al., 1998). Therefore, chromatin status is regulated through the recognition of histone modifications by chromatin remodelers or through the modification of histones by remodeler complexes, and open chromatin is necessary for active transcription. Once established, this accessible chromatin state permits the assembly of polymerase preinitiation complexes and subsequent transcriptional activity.

1.5.2 RNA polymerase II-mediated transcription

RNA polymerase II (RNA Pol II) consists of 12 subunits and transcribes protein-coding genes in eukaryotic cells (Armache et al., 2003). The C-terminal domain (CTD) of RNA Pol II contains tandem heptad repeats (consensus Tyr1-Ser2-Pro3-Thr4-Ser5-Pro6-Ser7) and the phosphorylation of Ser2 and Ser5 is critical for regulating progression through the transcriptional process (Eick and Geyer, 2013). Transcription by RNA Pol II comprises initiation, pausing, elongation and termination phases, where each phase is intricately regulated (Chen et al., 2018a; Harlen and Churchman, 2017). The initiation begins with the assembly of a preinitiation complex, which composes initiation factors and RNA Pol II at core promoters (Cheung and Cramer, 2012). DNA strands are then unwound by the complex to form a transcription bubble. Subsequently, the preinitiation complex phosphorylates Ser5 on RNA Pol II to allow RNA Pol II to engage in RNA synthesis (Figure 1.11).

Shortly after transcription initiation, RNA Pol II synthesises a short transcript and pauses at promoter-proximal regions (around 10 nucleotides downstream of the transcription start site). The pausing is finely regulated by protein complexes that establish or release paused RNA Pol II, including 5,6-Dichlorobenzimidazole 1-β-D-ribofuranoside (DRB)-sensitivity-inducing factor (DSIF), negative elongation factor (NELF), Pol-II-associated factor complex (PAFc) and positive transcription elongation factor b (P-TEFb) (Marshall and Price, 1995; Vos et al., 2018; Wada et al., 1998; Yamaguchi et al., 1999). The subsequent transition of RNA Pol II from initiation to productive elongation requires p-TEFb a CDK9/CycT1 driven kinase to phosphorylate the Pol II carboxy-terminal domain (CTD), DSIF and NELF to release RNA Pol II from the promoter-proximal site (Fujinaga et al., 2004; Marshall et al., 1996; Yamada et al., 2006) (Figure 1.11).

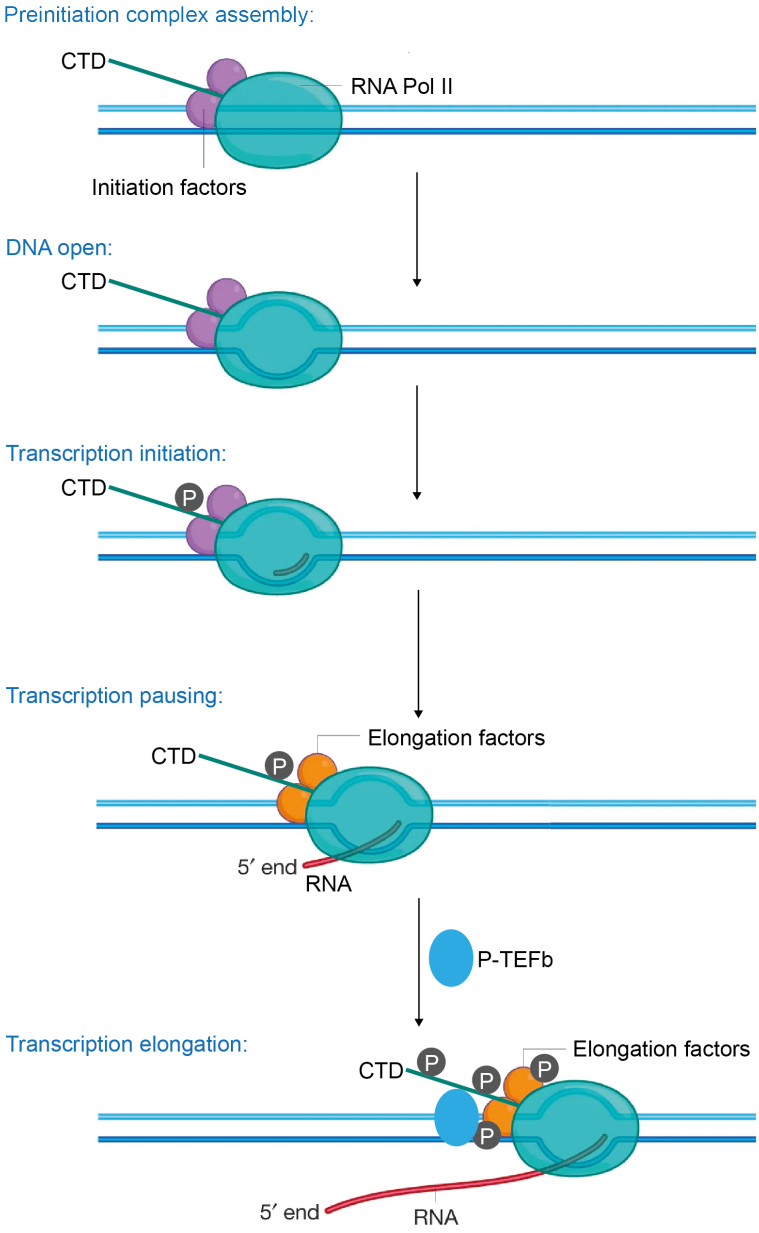

Figure 1.11: Key steps of RNA Pol II-mediated transcription.
RNA Pol II associates with initiation factors and assembles at promoter DNA to form a preinitiation complex. DNA is subsequently opened, exposing the DNA template strand. Serine 5 in RNA Pol II CTD gets phosphorylated and transcription initiates. When the RNA grows to a critical length, RNA Pol II is bound by elongation factors, DSIF and NELF, and transcription pauses. P-TEFb phosphorylates elongation factors and serine 2 in Pol II CTD and thus RNA Pol II engages into transcription elongation. Finally, polymerase dissociation from DNA and transcription terminates (not shown). Figure is adapted and modified from Cramer, 2019.

Transcription terminates when RNA synthesis is completed (usually signed as 3'-poly(A) sequence in human) and the nascent RNA is released from template DNA followed by the release of RNA Pol II. RNA Pol II termination is prompt through different pathways, and the poly(A)-dependent pathway is the best-studied pathway in human cells. It has been shown that poly(A) sequence is recognised and bound by human cleavage and polyadenylation specificity factor (CPSF) (Nag et al., 2007). The poly(A)-CPSF interaction induces transcription pausing, which further recruits the cleavage stimulatory factor (CstF) to the terminating sites (Nag et al., 2007). Recruited CstF stimulates the CPSF-mediated cleavage and consequently release of paused Pol II (Kuehner et al., 2011). Although much less is known about the termination machinery, it performs important functions in cells. For example, termination prevents RNA Pol II from interfering with downstream elements and promotes polymerase recycling (Rosonina et al., 2006). Interestingly, it has been reported that during the transcription of mitochondrial gene *CSB II*, the nascent RNA and non-template DNA strand forms an RNA-DNA hybrid G4 structure and promotes transcription termination (Wanrooij et al., 2012), suggesting that G4s may also be associated with transcription termination machinery.

1.6 Roles of DNA G4s in transcription

1.6.1 Regulation of transcription by G4 stabilisation

As described in section 1.4.1, Hänsel-Hertsch et al. have shown that endogenous G4 structures mark sites of active transcription (Hänsel-Hertsch et al., 2016). Moreover, enhanced G4 formation is associated with increased transcriptional activity. These findings suggest a positive correlation between G4 formation and transcription. However, early studies, which employed G4-stabilising ligands to investigate the relationship between G4s and transcription, suggested that stabilised G4s may play a role in inhibiting transcription. For example, cationic porphyrin TMPyP4 has been shown to stabilise an intramolecular G4 structure from the G-rich sequence upstream of *MYC* promoter (Siddiqui-Jain et al., 2002). Treatment of cells with TMPyP4 caused an overall reduction of *MYC* expression. Comparable to this observation, TMPyP4, which was also reported to stabilise a G4 structure at the *KRAS* promoter, strongly inhibited *KRAS* promoter activity (Cogoi and Xodo, 2006). Isoallozazine ligands were also used as G4-stabilising small molecules, and cells treated with these ligands

showed a 50% reduction in *KIT* expression (Bejugam et al., 2007). Furthermore, whole transcriptome RNA-seq revealed that treatment with CM03, which shows preferential binding to and stabilisation of G4 structures over duplex DNA *in vitro*, resulted in a transcriptome-wide downregulation of genes enriched in G4s in human PDAC cells (Marchetti et al., 2018). However, more explicit evidence is required to confirm that these small-molecule bind to folded G4s in cells.

Although G4 ligand-mediated G4 stabilisation is reported to result in downregulation of transcription, it is important to note that repression of transcription may not be directly caused by the formation of G4s. Illustrating this, treatment with the G4-stabilising ligand PDS induced DNA damage and promoted growth arrest in human cancer cells (Rodriguez et al., 2012). The study also showed that PDS treatment led to reduced expression of G4-containing genes. However, PDS-induced DNA damage rather than G4 structures may be the cause of altered gene expression in cells. It remains possible that previous studies which treated cells with G4 ligands for a long time may trap G4 structures, with indirect effects, such as DNA damage, on transcription. Therefore, a full understanding of the link between G4s and transcription requires that the potential consequences of indirect, network effects of ligand-mediated G4 stabilisation on transcription be considered.

1.6.2 Investigating the role of G4s in transcription elongation

During active transcription, RNA Pol II pauses just downstream of the transcription start site. In addition to the protein complexes that interact with RNA Pol II, local features of genome structure, such as G4s, may contribute to RNA Pol II pausing. Computational analysis of human genes showed preferential promoter-proximal RNA Pol II pausing in the presence of downstream G4 motifs on both template and non-template strands (Eddy et al., 2011), which suggests a potential role for gene body G4 structures in regulating gene expression in human cells.

Other studies have also investigated the interaction of gene body G4s with transcription elongation. For example, by introducing a reporter gene construct which contains a G4-forming sequence in its promoter in *Escherichia coli*, a substantial inhibitory effect was observed (Holder and Hartig, 2014). Similar inhibitory effects were also observed in human

embryonic kidney cells using a luciferase reporter assay, where a G4-forming sequence was engineered into the template strand downstream of the promoter in the plasmid (Agarwal et al., 2014). These findings are consistent in suggesting that G4 structures in the template strand impair RNA Pol II progression. Agarwal et al. also found that transcription was not impacted by placement of the G4-forming sequence on the non-template strand downstream of the promoter (Agarwal et al., 2014). In support of this observation, a genome-wide analysis revealed that human genes with a greater number of G4s, particularly those with G4s located on the non-template strand up to 500 base pairs downstream of the transcription start site, are associated with higher than average transcription levels and RNA Pol II occupancy (Du et al., 2008). This suggests that G4s on the non-template strand may maintain an open DNA state and thereby aid transcription reinitiation. In opposition to this model, another study reported that G-rich sequence in the non-template strand resulted in reduced *in vitro* transcription by T7 polymerase (Belotserkovskii et al., 2017). In this case however, impaired transcription was ascribed to R-loop formation (see section 1.6.3) rather than G4s themselves, as T7-mediated transcription did not change in G4-stabilising conditions when compared to G4-destabilisation conditions.

Overall, the precise relationship between DNA G4s and transcription remains unclear and the subject of much debate. Furthermore, most studies to date have used reporter assays or *in vitro* approaches to investigate the relationship, without considering chromatin context. Given recent findings from G4 ChIP-seq experiments that suggest G4 formation is tightly linked to chromatin context (Hänsel-Hertsch et al., 2016), it will be important in the future to develop systems that allow the link between G4s and transcription to be interrogated in the context of chromatin.

1.6.3 The interaction of G4s with torsional stress and R-loops in transcription

During transcription initiation, the RNA polymerase preinitiation complex (PIC) binds and unwinds the DNA duplex to promote the engagement of RNA Pol II in active transcription. It has been proposed that negative torsional stress and supercoiling induced by counter-rotation of the transcription machinery affects upstream DNA topology (Harada et al., 2001; Liu and Wang, 1987) and promotes melting of duplex DNA, which in turn favours the formation of G4 structures. Support for this model comes from a study that examined the *in*

vitro transcription of the *MYC* gene (Kouzine et al., 2004). Using this system, Kouzine et al. showed that transcribing RNA polymerases drive the supercoil-sensitive far upstream element (FUSE) of the human MYC gene into a single-stranded DNA conformation. During this process, they observed that FUSE adopts a non-B-DNA conformation.

R-loops are RNA-DNA hybrids in which nascent RNA produced by transcription anneals with its template DNA. Emerging evidence has suggested that R-loops have functions in biological processes, such as transcription and DNA repair (Sollier and Cimprich, 2015). Dysregulation of R-loops also causes DNA damage and genome instability (Santos-Pereira and Aguilera, 2015). Similar to G4 structures, DNA features that favour R-loop formation include GC-richness and negative torsional tension (Ginno et al., 2012). The R-loop landscape mapped by ChIP-seq using enzymatically dead RNase-H1 (which shows specificity for DNA-RNA hybrids) revealed that R-loops preferentially form in promoter regions that are likely to form G4s *in vitro* (Chen et al., 2017). The co-occurrence of G4 and R-loop was termed a 'G-loop', in which a G4 forms on the non-template strand and the R-loop on the template strand (Duquette et al., 2004). When transcribing a plasmid genome, which contains G-rich sequences downstream of the promoter, *in vitro* or in *Escherichia coli*, Duquette et al. observed stable formation of loop structures by electron microscopy (Duquette et al., 2004). By the addition of digoxygenin-UTP during transcription and gold beads coupled to anti-digoxygenin antibodies, the authors were able to detect RNA in this loop structure, suggesting that it indeed forms an R-loop. Cleavage of one strand of the loop by GQN1, an endonuclease that specifically cleaves G4 DNA in the 5' of the stacked G-tetrads (Sun et al., 2001), provided evidence for the existence of a G4. More recently, Magis and co-workers found that the induction of G4 formation using G4-stabilising ligand PDS caused increased R-loop formation and extension in cells (De Magis et al., 2019). It has also been suggested that the PDS-induced R-loops spreading is associated with the formation of nearby G4s, as the R-loop extension regions possesses more experimentally observed G4 structures on the displaced strand. Together these findings suggest the existence of G-loops during active transcription.

1.7 Aims of the thesis

It has been indicated that G4s are prevalent and enriched in human regulatory regions such as gene promoters (Chambers et al., 2015; Hänsel-Hertsch et al., 2016; Kudlicki, 2016; Varizhuk et al., 2017). Moreover, endogenous G4 structures in human chromatin, which are located predominantly in nucleosome-depleted regions, mark elevated levels of transcription with the increase of chromatin accessibility (Hänsel-Hertsch et al., 2016). Together, the evidence suggests that regulatory nucleosome-depleted chromatin and elevated transcription shape the endogenous human G4 DNA landscape.

A great deal of effort has focused on investigating the relationship between DNA G4s and transcription. However at the onset of this study, models and data linking G4s to transcription were largely based on computationally predicted G4 sequences and relied on systems that lack chromatin context. There exists the need to transition from *in vitro* systems in which isolated G4 structures are engineered into plasmid constructs, to cellular models that more accurately reflect the endogenous environment of G4s. Doing so will be critical for obtaining a mechanistic understanding of how G4s influence transcription in cells. More specifically, one important outstanding question that remains to be addressed is whether an open chromatin conformation is sufficient to induce G4 formation or whether active transcription is also required. The overarching aim of this thesis is to investigate the causal relationship between G4 formation, active transcription and chromatin structure, to further the understanding of the biological functions of promoter G4s.

Aim 1: Emerging evidence suggests that endogenous G4 structures are associated with active gene transcription (Hänsel-Hertsch et al., 2016). *In vitro* work has suggested that active transcription can promote the formation of G4s (Duquette et al., 2004; Kouzine et al., 2004), suggesting a potential functional relationship between the two. To explore this possibility, the first aim of this thesis is to investigate the causal relationship between G4s and active transcription through the use of chemical intervention on specific aspects of the transcriptional process in combination with genomics-based approaches.

Aim 2: Having explored the relationship between the formation of G4 structures and active transcription, Aim 2 moves towards a comprehensive understanding of the interaction of

promoter G4s with transcription and chromatin. Aim 2 is to characterise the contribution of chromatin state to G4 formation. Focusing specifically on promoter G4s, I aimed to test the hypothesis that G4 formation is governed by the local chromatin environment.

Aim 3: The third aim arose from observations made in Aim 1 and Aim2, that G4 formation is determined by chromatin structure but not transcriptional activity. To interrogate this relationship further, I ask whether promoter G4s play a role in regulating transcription, and explore the potential mechanisms that underlie this relationship.

Chapter 2

Response of endogenous G4s to transcription inhibition

2.1 Background

Genome-wide enrichment of G4s at promoter proximal regions was first identified by computational methods that predict the folding of G4s by analysing the underlying sequence (Eddy and Maizels, 2006; Huppert and Balasubramanian, 2005; Todd et al., 2005). Many such G4 motifs were subsequently verified to fold into G4 structures *in vitro*. However, determining whether G4s fold at these loci in cells was more challenging. This goal was helped by the advent of G4 structure-specific antibodies, such as Sty49, HF2, BG4, 1H6 and D1 (Biffi et al., 2013b; Fernando et al., 2008; Henderson et al., 2014; Liu et al., 2016; Schaffitzel et al., 2001) (see section 1.3.2). Among these, BG4, a recombinant single-chain variable fragment (scFv) antibody was isolated by phage-display against TERRA, an RNA telomeric G4 (Biffi et al., 2013b). BG4 was found to bind specifically to DNA and RNA G4s *in vitro* (Kd = 0.5 to 18 nM) without detectable binding to other DNA or RNA structures (Biffi et al., 2013b, 2013a). Remarkably, BG4 did not display preferential binding to a specific G4 conformation for the G4s tested, which enabled its use in detection of different G4s in a cellular context.

The first map reflecting the cellular human G4 landscape was produced using BG4 antibody-based chromatin immunoprecipitation (ChIP) coupled with high-throughput sequencing (G4 ChIP-seq) (Hänsel-Hertsch et al., 2016, 2018). Formation of G4 structures was detected within cross-linked chromatin at high resolution, thus identifying specific genomic loci that fold into G4s in cells. This approach interestingly revealed that endogenous G4 structures mapped predominantly to nucleosome-depleted regions and marked sites of active transcription. BG4 also co-localised with RNA Polymerase II (RNA Pol II) and trimethylated histone H3 Lys4 (H3K4me3), a histone modification that marks transcriptionally active euchromatin, as seen by immunofluorescent staining (Hänsel-Hertsch et al., 2016). Genes containing a G4 ChIP-seq peak in their promoter regions showed significantly higher transcription levels than those lacking a promoter G4 ChIP-seq peak. Notably, previously published ChIP-seq peaks for transcription factors (TFs) such as XPB and SP1 were also

found highly enriched in the G4 ChIP-seq data, suggesting interactions of these proteins with G4s (Hänsel-Hertsch et al., 2016).

Several independent studies indicated that G4 structures act as a binding site for TFs. For instance, SP1, a transcription factor ubiquitously involved in regulating house-keeping genes, was observed to specifically bind a G4 structure from the *KIT* promoter *in vitro* (Raiber et al., 2012). The binding of SP1 to G4 showed affinity comparable to its previously defined canonical duplex DNA binding site and mutations in G4 sequence resulted in reduced SP1 binding. Transcriptionally-associated helicases XPD and XPB were also shown to bind G4s using gel mobility shift assay, while XPD was also found to unwind G4 structures *in vitro* (Gray et al., 2014). In another study, G4 structure within the *KRAS* promoter region was found to be occupied by Myc-associated zinc finger (MAZ) and poly(ADP-ribose) polymerase 1 (PARP-1), both of which are crucial factors for maintaining cellular transcription (Cogoi et al., 2010). Similarly, G4 structures from the *MYC* promoter transfected into cells were found to bind HNRNPK and CNBP, indicating these proteins are directly recruited to *MYC* promoter via their interaction with G4s (Sengupta et al., 2019).

Further evidence for G4-TF interactions came from a quantitative protein-DNA binding assay. By using a *MYC* promoter G4-oligonucleotide as a bait for interacting proteins from nuclear extracts, Makowski and co-workers discovered a multitude of G4 interacting TFs using mass spectrometry (Makowski et al., 2018). These included SP1, as well as subunits of transcription factor IIH (i.e. XPB and XPD), which have previously been shown to bind to G4-folding oligonucleotides *in vitro* (Gray et al., 2014; Raiber et al., 2012). Furthermore, recent studies systematically assessed the genome-wide interaction between endogenous G4s and all known TF binding sites deposited on ENCODE database (Hou et al., 2019; Spiegel et al., 2021). This revealed that TFs, such as SP2, FUS and NRF1, are recruited to promoter G4s. *In vitro* binding assays suggested that the interaction of these proteins with folded G4s was comparable to their binding to the respective consensus duplex DNA targets. Notably, not only did these TFs bind to thousands of G4 sites in cells, but these interactions could also be interrupted via the use of G4 stabilising ligands, which further suggests direct binding to G4 structures in chromatin.

The evidence I have described so far has suggested that G4 structures are involved in active transcription. However, these observations rely on *in vitro* systems or on genomic correlations of G4s with sites of active transcription in cells. The causal relationship between G4s and the transcription process still needs to be established. Some indications towards such a functional relationship have been reported (see section 1.6.3). For example, negative torsional stress caused by RNA polymerase preinitiation complex promotes DNA duplex melting to form a single-stranded transcription bubble (Kouzine et al., 2004). This is hypothesised to promote the formation of transcription associated G4s. Furthermore, nascent RNA produced during transcription may anneal to its template DNA, forming an RNA-DNA hybrid known as R-loop. R-loops appear to form predominately in GC-rich non-template DNA strand at the promoter proximal region (Ginno et al., 2012). The formation of R-loops during active transcription could also promote the formation of G4 structures. This co-occurrence of G4 and R-loop was termed 'G-loop', where G4 forms on non-template strand and R-loop on template strand. The presence of G-loop formation was observed co-transcriptionally *in vitro* or in *E.coli* and visualised by electron microscopy (Duquette et al., 2004). In agreement with this, a newly developed ChIP-seq assay based on enzymatically dead RNase-H1 mapped R-loops revealed that R-loops are enriched at the promoter regions which have the potential to form G4 structures (Chen et al., 2017).

In spite of these studies, we still lack a clear understanding of the causal association between G4 formation and active transcription. In other words, is active transcription necessary for the formation and maintenance of G4s, in particular promoter G4s. To systematically explore the relationship between transcription and promoter G4s, I employed K562, human chronic myelogenous leukemia cells which have been extensively characterised by ENCODE (Gerstein et al., 2012). To examine whether active transcription promotes G4 formation or G4 formation enhances transcription, I used small molecules 5,6-dichlorobenzimidazole1-b-D-ribofuranoside (DRB) and triptolide to inhibit RNA Pol II mediated transcription elongation and initiation respectively, and then assessed the changes in endogenous G4 landscape, particularly promoter G4s, by G4 ChIP-seq (Figure 2.1). This enabled us to investigate how G4s respond to altered RNA Pol II activity in cells.

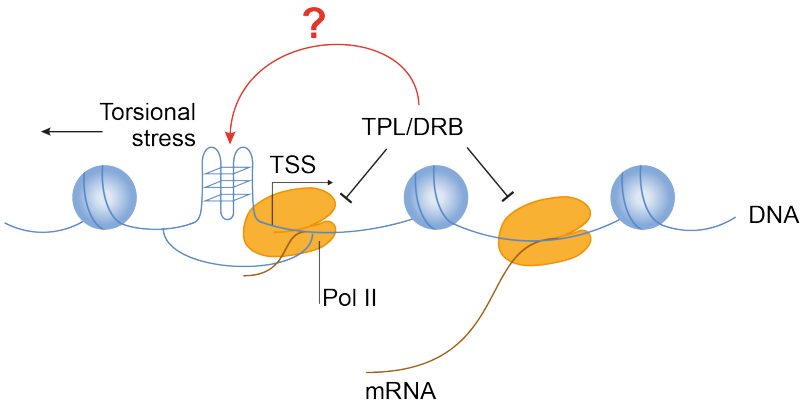

Figure 2.1: Investigating the response of endogenous G4s to transcriptional inhibition.
Graphical representation of experimental design examining how endogenous G4s respond to transcriptional inhibition. Active transcription induced torsional stress is hypothesised to promoter G4 folding. To examine whether the formation of G4s is promoted by active transcription, transcription elongation and initiation are inhibited by DRB and TPL respectively, and the change of G4 landscape is assessed by G4 ChIP-seq.

2.2 Objectives

The aim of this chapter is to understand the mechanistic interaction between active transcription and the formation of G4s. The first objective is to purify and validate the BG4 antibody, and to ascertain its purity and affinity for G4s *in vitro*. The second objective is to explore the correlation between G4s and transcription in K562 cells by generating a map of endogenous G4s using G4 ChIP-seq. The final objective is to study the response of endogenous G4s following transcriptional inhibition using DRB and TPL.

2.3 Results

2.3.1 Preparation and validation of BG4 antibody

G4 ChIP-seq was employed to map endogenous G4s in cells throughout this study. This technique relies on the G4 structure-specific single chain antibody BG4 (Biffi et al., 2013b), which I first expressed in *Escherichia coli* (*E. coli*), purified and validated. BG4 has 287 amino acids, a predicted molecular weight (MW) of 30.68 kDa and is dual-tagged with FLAG and poly-histidine (Figure 2.2A). The quality of recombinant BG4 was first assessed using a sodium dodecyl sulphate (SDS)-polyacrylamide gel electrophoresis (PAGE) protein analysis followed with Coomassie staining (Figure 2.2B). Recombinant BG4 demonstrated one major band sized around 31 kDa as anticipated. Smaller amounts of contaminating proteins sized 70kD, 25kD and 10kD were also observed. Since BG4 was expressed in *E. coli* derivative BL21 (DE3), a histidine-rich strain, these contaminants are likely to be arnA (70kD), SlyD (27kD) and the smaller proteins that are often co-purified with His-tagged target proteins (Andersen et al., 2013; Robichon et al., 2011).

A titration of bovine serum albumin (BSA) was used as protein standards to allow determination of BG4 concentration. To measure the concentration of BG4 antibody extracted from *E. coli*, a standard curve was generated by plotting the band intensity of the BSA standards against the known BSA concentration (Figure 2.2C). The curve was fitted using the linear least square method and the equation of the curve was calculated. By fitting the BG4 band intensity into standard curve, BG4 concentration was estimated to be 4.2 μM.

A

BG4 protein sequence:

MAEVQLVQSGAEVKKPGASVKVSCKASGYTFTSYSISWVRQAPGQGLEWMGWISAYNG
NTSYAQKLQGRVTITADKSTSTAYMELSSLRSEDTAVYYCAKAGHRSGRYNNWFDPWGQ
GTLVTVSSLE**GGGGSGGGGSGGGAS**QSELTQPPSVSVAPGQTARITCGENNIGSKNVHW
YQQKPGQAPVLIIYRGSNRPSGIPERFSGSNSGNTATLTISRAQAGDEADYYCQVFDRRSD
HPVVFGGGTKLTVLGAAASA**HHHHHH**KL**DYKDHDGDYKDHDIDYKDDDDK***

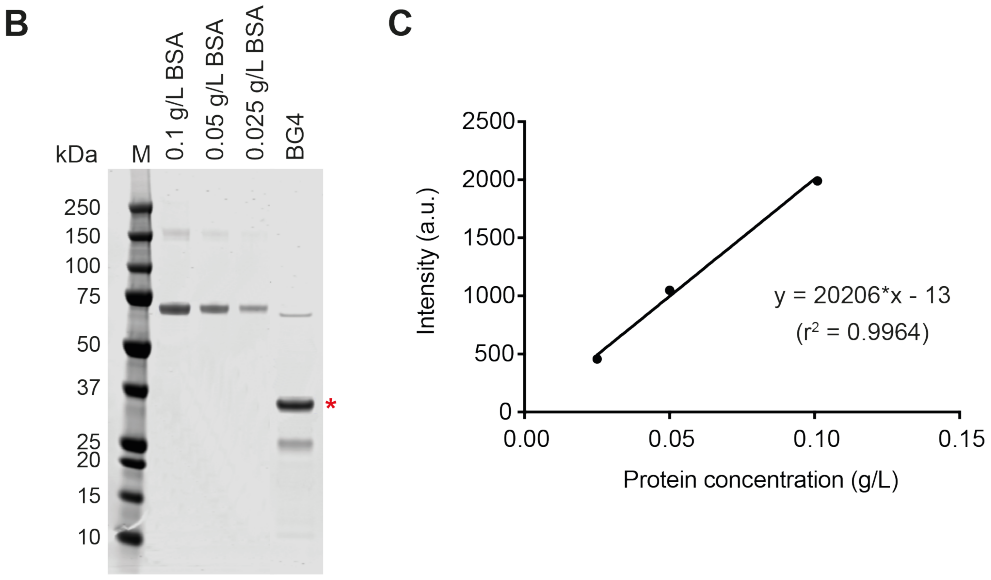

Figure 2.2: Validation of BG4 antibody.
(A) Protein sequence of BG4 with 6 His-tag, 3x FLAG and linker highlighted in green, blue and red respectively. **(B)** Coomassie blue staining of purified FLAG-tagged BG4 antibody and BSA standards separated by SDS PAGE. First lane contained molecular weight marker with proteins of denoted molecular weight. BG4 antibody is indicated with a red asterisk. **(C)** Protein standard curve generated using densitometry analysis performed on the BSA standards at the following concentrations: 0.025, 0.05 and 0.10 g/L. (a.u. = arbitrary unit).

Enzyme-linked immunosorbent assay (ELISA) was then performed to evaluate the affinity and specificity of purified BG4 against oligonucleotides of folded G4 structures. The G4 motif at the *MYC* promoter is able to form stable DNA G-quadruplex (Ambrus et al., 2005). G to C mutations within the central tetrad ablate the formation of the G4 structure and this provides a negative control for BG4 binding. The presence and absence of G4 structure in respective *MYC* and mutated *MYC* oligonucleotides were confirmed by circular dichroism (CD) spectroscopy. G4 structures depict a circular ellipticity maxima at 265 nm and minima at 245 nm (del Villar-Guerra et al., 2018). This pattern is altered upon mutation of the central stack of guanines, which indicates disruption of the G4 structure (Figure 2.3A).

Binding curves determined by ELISA showed that BG4 bound to *MYC* G4 oligonucleotide with low nanomolar affinity (K_d = 4.8 ± 0.4 nM) with negligible binding to mutated MYC oligonucleotide (Figure 2.3B), which is comparable to BG4 binding affinity previously described (K_d = 1.5 ± 0.2 nM) (Biffi et al., 2013b). The *E. coli* contaminants observed in purified BG4 did not markedly alter the BG4 affinity to either G4 structures or non-G4-forming oligos as seen by literature-matched K_d values. Additionally, these contaminants are unlikely to interfere with G4 ChIP-seq since they lack the FLAG tag that is utilised during this protocol. These results suggest that BG4 has been successfully generated with appropriate activities, which can be used for the following experiments.

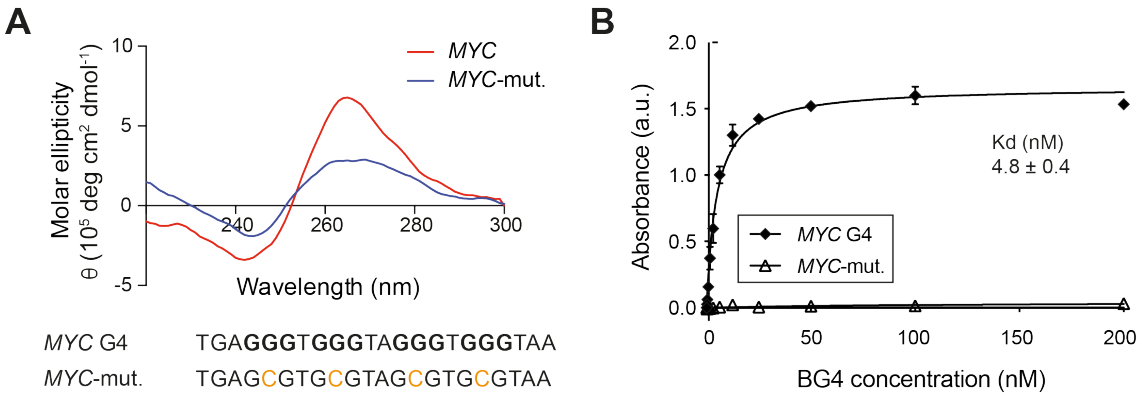

Figure 2.3: BG4 selectively binds to MYC G-quadruplex (G4).
(A) Circular dichroism spectra of *MYC* G4-folding oligonucleotides (red) and *MYC*-mut, oligonucleotides that are unable to fold into a G4 (blue). Sequences of *MYC* oligonucleotides are indicated and mutations were highlighted in yellow. **(B)** Binding affinity measurements performed by ELISA show the binding of BG4 antibody to *MYC* G4 structure and absence of binding to non-G4-forming control oligonucleotides (*MYC*-mut). BG4 binding curves were fitted using nonlinear regression analysis. Calculated dissociation constants (K_d) are indicated. Error bars represent the s.e.m. (n=3). (a.u. = arbitrary unit)

2.3.2 Characterisation of endogenous G4s in K562 cells

Originally, Hänsel-Hertsch et al. mapped endogenous G4 structures within chromatin of HaCaT cells using G4 ChIP-seq (Hänsel-Hertsch et al., 2016). G4s were found predominantly in nucleosome-depleted regions of the genome. Moreover, endogenous G4s were enriched in regulatory regions such as promoters and 5' UTRs. It was next observed that the G4 landscape in chromatin differs between cell types (Hänsel-Hertsch et al., 2018). Since K562 cells were employed as the model system throughout this study, I first characterised the G4 landscape with G4 ChIP-seq in K562 cells. Three biological replicates and three technical G4 ChIP replicates per biological replicates were performed. Since G4 ChIP-seq generally demonstrates lower signal-to-noise ratio when compared to other ChIP-seq approaches, so technical replicates help to assess reproducible signals.

Prior to sequencing, G4-ChIP efficiency was measured using quantitative PCR (ChIP-qPCR) for previously defined positive and negative control regions (Hänsel-Hertsch et al., 2018). Primers flanking G4 ChIP peak observed upstream of the *RPA3* gene and a G4 free region upstream of the *TMCC1* gene were used for amplification. The recovery fractions (referred

to as % input) were calculated by normalising the amplification results (i.e. the quantification cycle value, C_q) of immunoprecipitated DNA to those of input DNA. In addition, fold enrichment of *RPA3* signal over *TMCC1* signal was determined to assess G4 ChIP signal-to-noise ratio. ChIP-qPCR results showed that recovery fractions at *RPA3* were more than 5% and at *TMCC1* less than 0.5% for all biological replicates (Figure 2.4A). BG4 immunoprecipitated DNA was enriched 20-fold more at *RPA3* than at the negative control *TMCC1* (Figure 2.4B). The G4 ChIP is considered to be of acceptable quality when *RPA3* yields DNA-input recovery ≥ 5% and the relative fold enrichment is over 5-fold (Hänsel-Hertsch et al., 2018). Thus, ChIP-qPCR results suggested sufficient quality for subsequent sequencing.

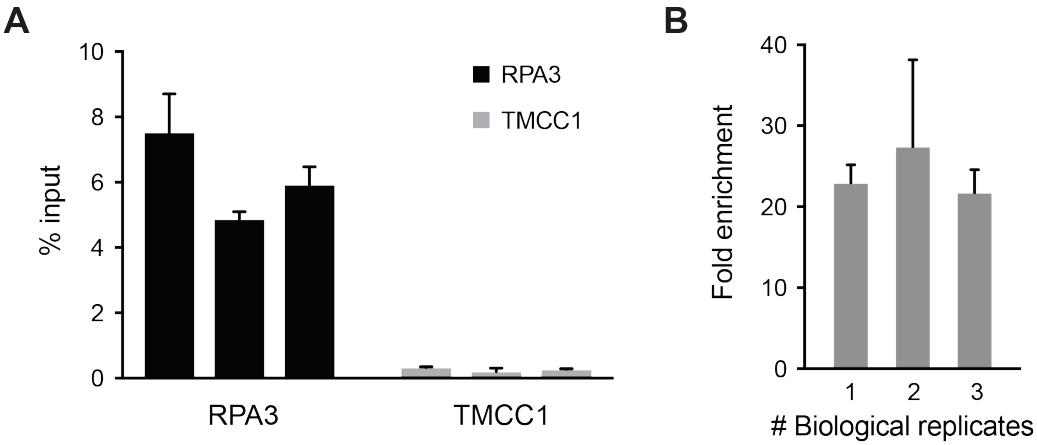

Figure 2.4: G4 ChIP-qPCR for selected G4 ChIP peaks in K562 cells.
(A) The recovery ratio of regions of interest determined by ChIP-qPCR. RPA3 region or TMCC1 region was used as a positive control or a negative control respectively.The ratio is normalised by input genomic DNA. **(B)** G4 ChIP signal-to-noise ratio determined by fold enrichment. Error bars represent the s.d. (n=3).

Sequencing reads from each biological replicate were normalised by counts per million and plotted against each other on a log2-scale scatter plot (Figure 2.5). The Pearson's correlation coefficient (r), which measures the strength of the association between biological replicates, demonstrated good reproducibility of the G4 ChIP-seq signal (r > 0.96). A consensus set of G4s was then defined as G4 ChIP-seq peaks present in at least two out of three biological replicates. In K562 cells, 19,048 consensus G4s were identified by G4 ChIP-seq (Figure 2.6). As was observed previously for HaCaT cells (Hänsel-Hertsch et al., 2016), the majority of the consensus G4 peaks (14,496/19,048, 76.1%) correspond to G-quadruplexes identified *in vitro* as polymerase stall sites in purified human genomic DNA by G4-seq (Chambers et al., 2015) (Figure 2.6A).

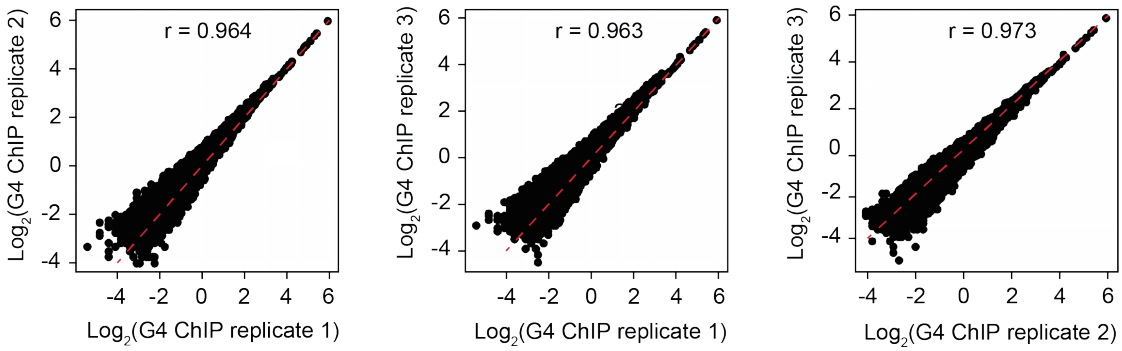

Figure 2.5: Biological replicates of G4 ChIP-seq are highly reproducible.
Scatter plots (read counts per million in log2 scale) showing the correlation of G4 ChIP-seq reads between biological replicates in K562 cells. Pearson's correlation coefficient (r) is shown in each diagram.

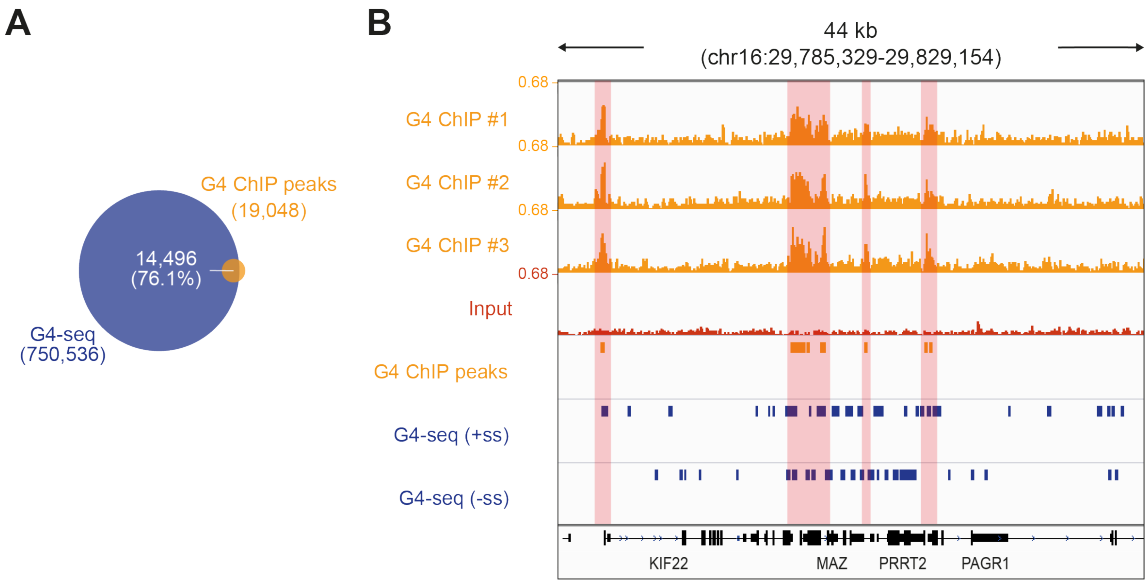

Figure 2.6: G4 ChIP peaks in K562 cells predominantly overlap with motifs that can form G4s *in vitro*.
(A) Venn diagram indicate positional overlap of G4 ChIP peaks and motifs that form G4s *in vitro* (G4-seq). **(B)** Genome browser view of example genes in K562 cells. Tracks from top to bottom are G4 ChIP-seq (yellow), input (red) and consensus G4 ChIP peaks (yellow) for K562 cells, and *in vitro* forming G4s (G4-seq) (blue) on the forward (+ss) and reverse (-ss) strand. G4 ChIP-seq peaks are masked in red.

As highlighted by G4-seq, canonical G4 structures, defined as $G_{\geq 3}N_{1-7}G_{\geq 3}N_{1-7}G_{\geq 3}N_{1-7}G_{\geq 3}$, are not the only G4s that can fold *in vitro* (Chambers et al., 2015). Non-canonical G4s including those with long loops and single nucleotide bulges were also observed as folded in the human genome. Furthermore, both canonical and non-canonical G4 structures were identified by G4 ChIP-seq in HaCaT cells (Hänsel-Hertsch et al., 2016). Motif analysis of consensus G4 peaks from K562 cells showed that only 25% of sites identified by BG4 comprised a canonical G4 motif (Figure 2.7A). A large number of G4 peaks consisted of non-canonical G4 motifs comprising long loops (18.1%), bulges (16.6%) and 2-tetrad (25.3%) G4 structures. The remaining peaks (in the 'other' category) may be structures containing multinucleotide bulges or topologies comprising both long loops and bulges, which are difficult to account for computationally.

To test if the occurrence of these motifs in G4 ChIP peaks is statistically significant when compared to random chance, peaks were randomly shuffled across the genome and the

ratio of G4 motifs in consensus G4 peaks to randomly shuffled peaks was calculated (referred to as fold enrichment). For all canonical and non-canonical G4 motifs, except two-tetrad G4s, this ratio was found to be greater than one (Figure 2.7B), indicating that the likelihood of detecting these motifs occurring within the G4 ChIP-seq peaks is unlikely to be by chance. The higher the fold enrichment value, the greater is the likelihood that the G4 ChIP-seq peak is bona fide. Three tetrad G4s with short loops of 1-3 nucleotides showed the greatest fold enrichment (~20-fold), whereas increasing the loop length or the presence of bulges resulted in incremental reduction in fold enrichment. Two-tetrads G4 motifs are very abundant within the genome and therefore were found not to be significantly enriched within the consensus G4 peaks.

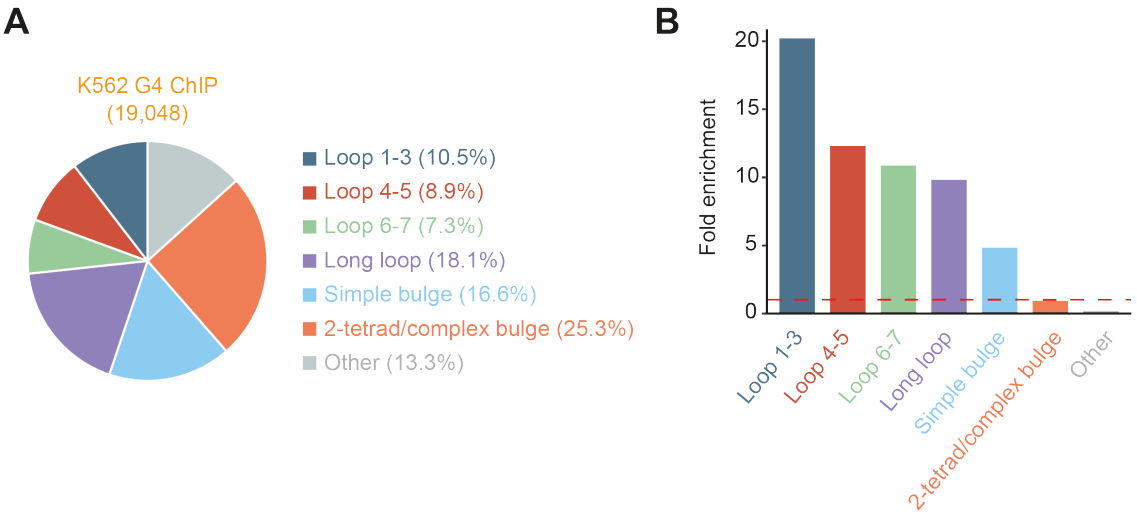

Figure 2.7: Motif analysis of consensus G4 peaks identified using ChIP-seq in K562 cells.
(A) Proportion of each structural class of G4 motif as shown in pie chart. Loop size 1-3, 4-5 and 6-7 represent G4s have at least one loop of indicated length. Long loop comprises G4 with loops longer than 7nt, whereas simple bulge represents G4 structure containing a 1-7 base bulge within one of its G-stretch. 2-tetrads/complex bulge comprise G4 structures with only 2 G-tetrads stacked together or containing more than one bulges, and other indicates G4 structures that do not fall into any of the former categories. (B) Fold enrichment over random for each G4 motif.

I then assessed the distribution of consensus G4 peaks within annotated genomic features (Figure 2.8). Endogenous G4s in K562 cells were found highly enriched in functional regions, showing the greatest fold enrichment within promoters (defined as TSS ± 500 bp; ~24 fold). These findings agree with earlier studies (Chambers et al., 2015; Hänsel-Hertsch et al., 2016; Huppert and Balasubramanian, 2007) and strongly suggest conserved relationship between G4 formation and regulation of transcription. The next most enriched category of genomic features in consensus G4 peaks was 5' UTRs showing 20-fold enrichment, although only a very small proportion of peaks (~2%) were found within 5' UTRs. Despite 10% of consensus G4 ChIP peaks being identified within introns and intergenic regions, these genomic features were not found to be enriched.

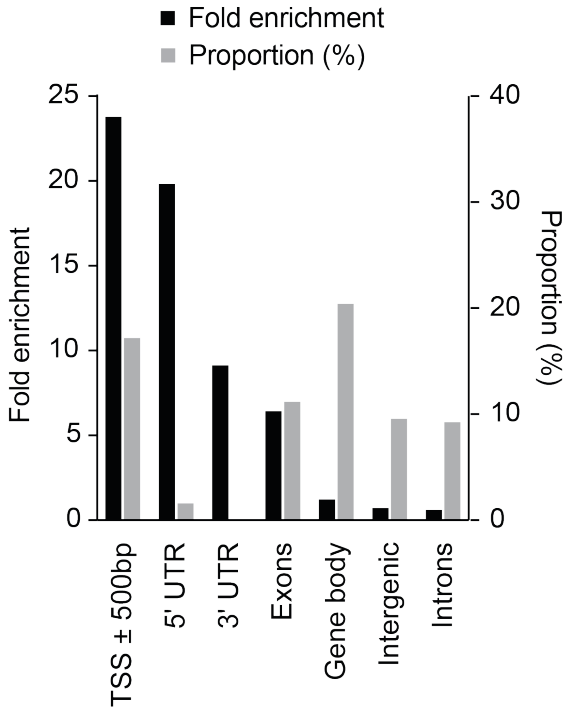

Figure 2.8: G4s in K562 cells are enriched within functional genomic regions.
Fold enrichment of consensus G4 peaks compared to randomly shuffled G4 peaks in regions that have the potential to form G4s *in vitro* (G4-seq) and the proportion of G4 peaks within different genomic features. TSS, transcription start site. UTR, untranslated regions.

Since endogenous G4s are found in nucleosome depleted regions (NDRs) (Hänsel-Hertsch et al., 2016), I next established a map of chromatin accessibility for K562 cells to enable the comparison with G4s. I performed the assay for transposase-accessible chromatin using sequencing (ATAC-seq) to study the chromatin accessibility in K562 chromatin. The original protocol employs hyperactive transposase Tn5 to directly digest cell lysates and captures the accessible chromatin regions for sequencing (Buenrostro et al., 2013, 2015). However, this protocol produces sequencing libraries with a relatively high mitochondrial DNA (mtDNA) content with over 50% of reads mapping to mtDNA, which severely diminishes the sequencing depth over nuclear chromosomes. In a modified ATAC-seq protocol (Calviello et al., 2019), digestion by Tn5 is performed on whole cells with dramatically decreases mtDNA contamination, with only 18% of reads mapped to mtDNA. It is reasoned that by avoiding detergent lysis, mitochondrial membranes stay intact while the nuclear membrane remains permeable. I first compared the two ATAC-seq protocols in K562 cells and assessed the quality of data obtained. Indeed, the modified version of the protocol resulted in 7-fold less reads mapping to mtDNA (21% vs. 3%) compared to the original protocol. In addition, the duplication rate in original protocol was higher than the modified one. Figure 2.9 shows the corresponding fragment size distributions from two different ATAC-seq protocols. The original protocol (left panel) did not show a clear distribution of mono-nucleosomal or di-nucleosomal DNA, whereas the modified ATAC-seq (right panel) showed clear nucleosomal fragment distribution as previously observed (Yan et al., 2020), suggesting higher quality ATAC-seq libraries. I therefore used the modified ATAC-seq protocol (hereafter defined as ATAC-seq) for the following studies for assessing genome-wide chromatin accessibility.

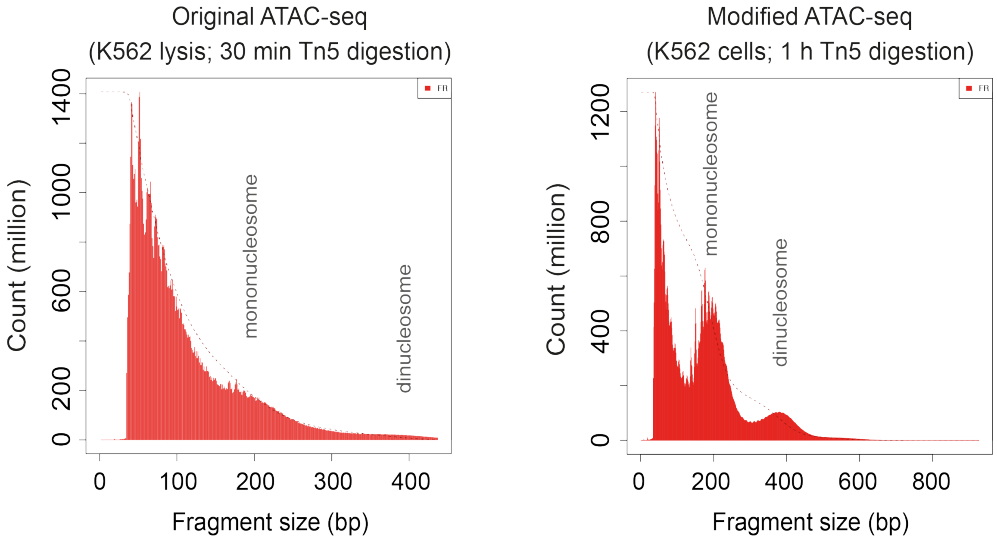

Figure 2.9: Fragment size distribution of two different ATAC-seq protocols in K562 cells.
Fragment size distribution for two different protocols, original and modified ATAC-seq in K562 cells. The nucleosomal structure patterns are as indicated (see *ref. Yan et al., 2020*).

ATAC-seq was performed on three independent biological replicates and sequencing data demonstrated very good correlation (Pearson's correlation r > 0.96; Figure 2.10). ATAC-seq peaks present in at least two out of three biological replicates were considered consensus. Since optimised ATAC-seq shows less mtDNA contaminants and increased sequencing depth over nuclear chromosomes, more than 40,000 consensus peaks were identified as accessible chromatin regions in K562 cells, whereas ~23,000 peaks were reported in previously published data (Liu et al., 2017) (Figure 2.11A). The majority of the published ATAC-seq peaks (19,800/23,028, 86%; GEO accession no. GSE99173) were identified by ATAC-seq I performed in K562 cells. Furthermore, regions immediately upstream of TSS demonstrated increased chromatin accessibility as exemplified in genome browser view in Figure 2.11B. I then investigated the relationship between endogenous G4s in K562 chromatin as defined by G4 ChIP-seq with sites of open chromatin as defined by ATAC-seq. Consensus G4 peaks were found to be predominantly located at sites of open chromatin, with the majority of G4 peaks (88.2%) overlapping a subset of ATAC-seq peaks in K562 cells (Figure 2.12). These data in K562 cells are in agreement with the G4 landscape in HaCaT cells and further confirm the idea that G4s primarily form within the NDRs upstream of TSSs.

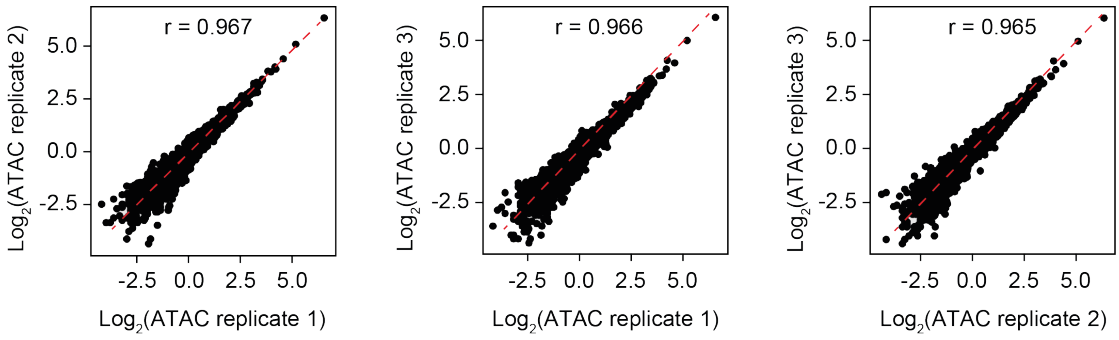

Figure 2.10: Three biological replicates of ATAC-seq in K562 cells are highly reproducible.
Scatter plots (read counts per million in log2 scale) showing the correlation of genome-wide ATAC-seq signal between biological replicates in K562 cells. Pearson's correlation coefficient (r) is shown in each diagram.

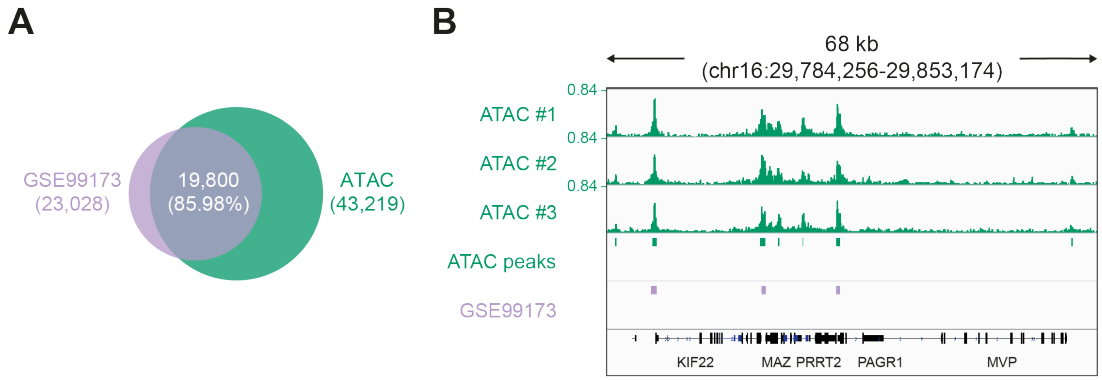

Figure 2.11: ATAC-seq data compare well with published data.
(A) Venn diagram showing the overlap of open chromatin regions (i.e. peaks) identified by ATAC-seq in K562 cells by this study and previously published ATAC-seq data (GSE99173). **(B)** Example of genome browser view of open chromatin sites in K562 cells. Tracks from top to bottom are three biological replicates of ATAC seq (green), consensus ATAC peaks (green) and published data from ENCODE describing open chromatin regions (purple).

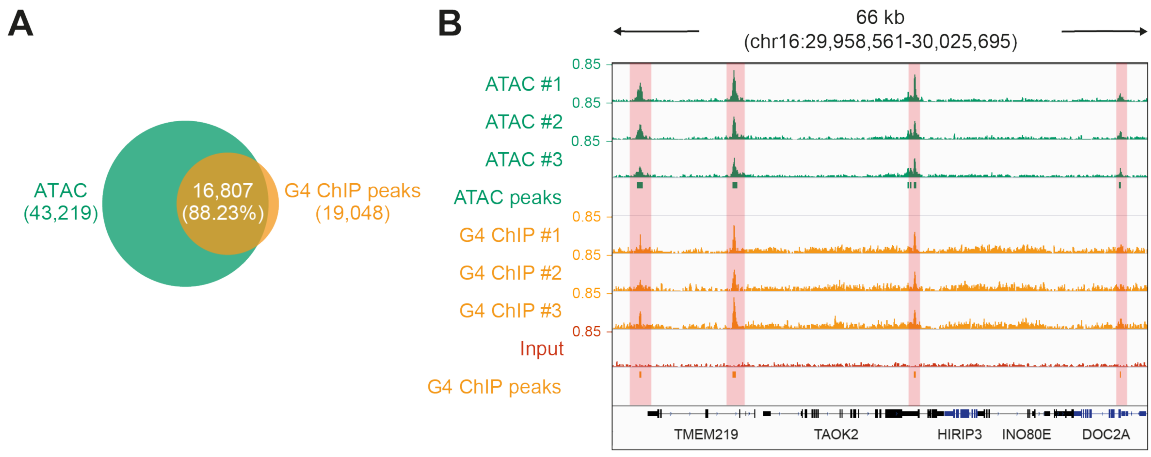

Figure 2.12: G4 ChIP peaks in K562 cells are predominantly located in open chromatin regions.
(A) Overlap between G4 ChIP peaks and ATAC-seq peaks. **(B)** Example genome browser view of G4 ChIP-seq and ATAC seq in K562 cells showing co-localisation (red-masked) of NDRs and G4s in the upstream region of genes. Tracks from top to bottom are three biological replicates of ATAC-seq (green), consensus ATAC peaks (green), three biological replicates of G4 ChIP-seq (yellow), input for G4 ChIP-seq (red) and consensus G4 ChIP peaks (yellow) in K562 cells. Genomic coordinates indicate track range.

2.3.3 G4 structures mark promoters with increased RNA polymerase II occupancy and enhanced transcription

Since consensus G4s in K562 chromatin are predominantly enriched in promoters (Figure 2.8), I further examined the relationship between promoter G4s and transcriptioal activity. First, I used the published RNA-seq dataset to look at steady-state transcript levels (GEO accession no. GSE88473; Dunham et al., 2012). Genes marked by a promoter G4 peak were found to show significantly higher expression (log_2TPM at non-G4-marked promoter = 2.2, log_2TPM at G4-marked promoter = 3.8; $p < 2.2 \times 10^{-16}$) as measured by RNA-seq when compared to those lacking a G4 peak but containing a promoter G4 motif that forms *in vitro* (G4-seq; Figure 2.13). These findings are in agreement with previous studies within HaCaT cells and primary keratinocytes where promoter marked by G4s shows higher transcriptional levels (Hänsel-Hertsch et al., 2016).

In the eukaryotic genome, RNA Pol II is recruited to the promoters of coding genes and higher levels of RNA Pol II occupancy at the TSS are associated with increased gene

expression (Cho et al., 2016; Quinodoz et al., 2014). Since the presence of promoter G4s positively correlate with enhanced gene expression, I reasoned that this would also be reflected in elevated RNA Pol II occupancy. To test this, ChIP-seq for RNA Pol II was performed on five independent biological replicates of K562 cells. The data obtained showed high reproducibility as seen by the Pearson's correlation coefficient r > 0.97 between individual biological replicates (Figure 2.14). Promoters marked with an endogenous G4 showed significantly higher RNA Pol II occupancy compared to their non-G4 marked counterparts that had promoters with open chromatin and G4 motifs that form G4 *in vitro* identified by G4-seq (log_2 Pol II at non-G4-marked promoter = 2.1, log_2 Pol II at G4-marked promoter = 3.6; $p < 2.2 \times 10^{-16}$; Figure 2.15). The presence of a promoter G4 thus not only correlates with elevated gene expression but also increased RNA Pol II occupancy. To further dissect the relationship between G4 formation, transcription and chromatin, the remainder of the study will mainly focus on promoter G4s (TSS ± 500 bp) of actively transcribed genes as defined by their RNA Pol II occupancy.

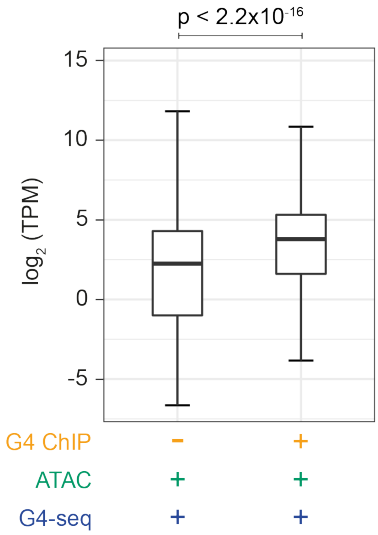

Figure 2.13: Endogenous G4 formation is correlated with elevated gene expression in K562 cells.
Box plot for the distribution of gene expression levels (displayed in transcripts per million (TPM), log2 scale) at promoters that feature an ATAC-seq peak and a G4 motif that form G4 *in vitro* (G4-seq) in comparison to promoters that have a G4 ChIP-seq peak, ATAC-seq peak and G4 motif that form G4 *in vitro* (G4-seq) in K562 cells. Wilcoxon test: $p < 2.2 \times 10^{-16}$.

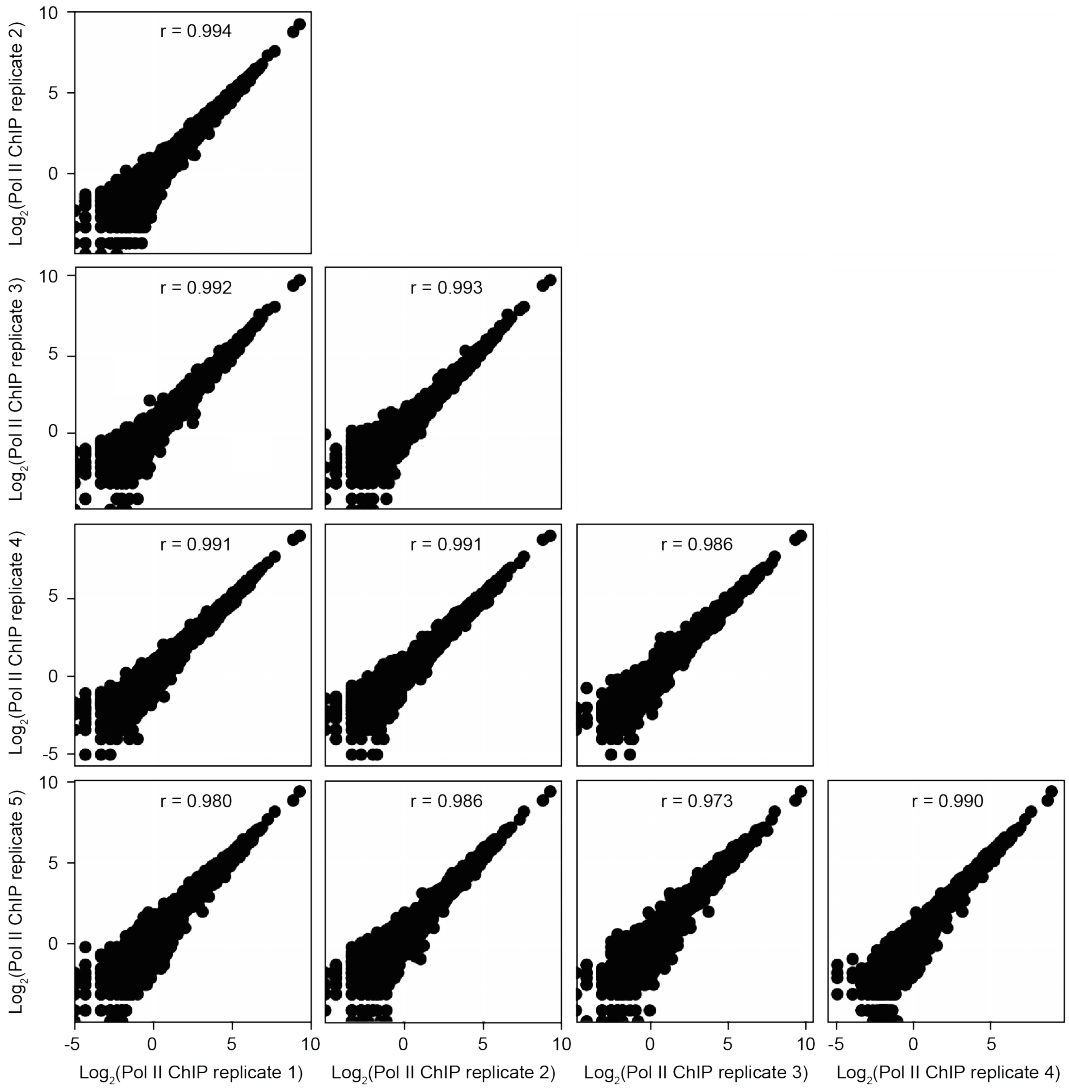

Figure 2.14: Five biological replicates of RNA Pol II ChIP-seq show high reproducibility.
Scatter plots (read counts per million in log2 scale) showing the correlation of genome-wide RNA Pol II ChIP-seq signal between biological replicates in K562 cells. Pearson's correlation coefficient (r) is shown in each diagram.

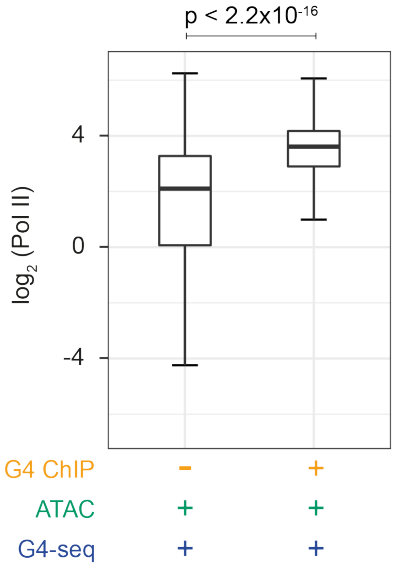

Figure 2.15: Endogenous promoter G4 formation correlates with elevated RNA Pol II occupancy in K562 cells.
Box plot of RNA Pol II ChIP-seq signal (displayed in read counts per million (CPM), log2 scale) at promoters in accessible chromatin (ATAC +) with or without an endogenous G4 (G4 ChIP + and G4 ChIP -) but having sequence motifs that can fold into G4 structures *in vitro* (G4-seq) in K562 cells. Wilcoxon test: $p < 2.2 \times 10^{-16}$.

2.3.4 5,6-Dichlorobenzimidazole 1-β-D-ribofuranoside (DRB) inhibits transcription elongation in K562 cells

Having established that the formation of endogenous G4 structures is correlated with elevated RNA Pol II occupancy and increased expression in K562 cells, the next aim was to evaluate whether active transcription is required for the formation of promoter G4s. To test this, I asked whether inhibition of transcription by RNA Pol II leads to loss of G4s. As described in Chapter 1, RNA Pol II mediated transcription is an intricately regulated process which comprises initiation, pausing, elongation and termination phases (Chen et al., 2018a; Harlen and Churchman, 2017). Since G4s are highly enriched at promoters where transcription initiates and productive elongation starts, the impact of inhibition of transcription initiation and elongation upon G4 landscape was investigated.

As described previously (see section 1.5.2), P-TEFb is critical for the transition of RNA Pol II from initiation to productive elongation (Price, 2000). P-TEFb is a CDK9 driven kinase that

phosphorylates the carboxy-terminal domain (CTD) of RNA Pol II, DSIF and NELF to promote productive elongation (Fujinaga et al., 2004; Marshall et al., 1996; Yamada et al., 2006). To perturb transcription elongation, I treated K562 cells with the CDK9 inhibitor DRB (Baumli et al., 2010; Laitem et al., 2015). Cells were treated for 1 h with a standard concentration of DRB (100 µM) that has previously been reported to cause depletion of elongating polymerase as seen by Pol II ChIP-sequencing (Erickson et al., 2018). To assess the efficiency of DRB treatment, cell lysates were analysed using SDS-PAGE and western blotting. The level of serine 2 phosphorylation of RNA Pol II CTD (Pol II Ser2-P) was found to be substantially reduced following DRB treatment (Figure 2.16). Pol II Ser2-P marks productive transcription elongation (Egloff et al., 2012; Eick and Geyer, 2013), the reduction of Pol II Ser2-P thus signifies the successful inhibition of transcription elongation. Meanwhile, total RNA Pol II protein levels did not show significant changes, suggesting that transcription elongation inhibition by DRB did not affect overall RNA Pol II levels.

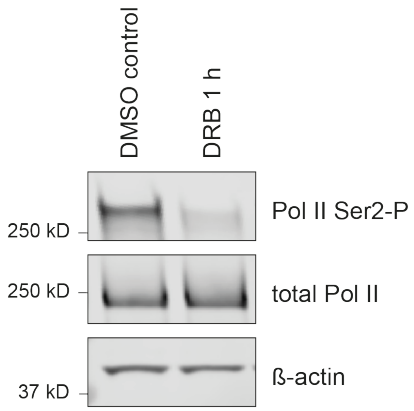

Figure 2.16: 5,6-Dichlorobenzimidazole 1-ß-D-ribofuranoside (DRB) inhibits transcription elongation.
Western blotting with antibodies against phosphorylated RNA Pol II Ser2 (Pol II Ser2-P), total RNA Pol II and ß-actin in whole cell lysates from K562 cells treated with or without 100 µM DRB for 1 h. ß-actin provides a loading control. Protein standards were used as molecular weight markers to determine protein size.

2.3.5 Inhibition of transcriptional elongation by DRB does not alter G4 landscape

I then performed G4 ChIP-seq after DRB-mediated transcription elongation inhibition to assess any effects on G4 formation. G4 ChIP-seq was performed on four independent biological replicates treated with 100 µM of DRB in DMSO for 1 h or with DMSO control. This identified 7,154 genome-wide consensus G4 peaks (i.e. G4 ChIP-seq peaks present in more than two biological replicates.) in DMSO-treated K562 cells (Figure 2.17). These peaks showed similar genomic distribution to consensus G4 peaks described in section 2.3.2 (Figure 2.7 and Figure 2.8). The majority of the G4 peaks identified in DMSO-treated cells overlapped with G4 peaks in DRB-treated K562 cells (6,594/7,154; 92.2%). A 20% increase (8,780 peaks vs. 7,154 peaks) in the total number of consensus G4 ChIP peaks was observed upon DRB treatment. However, peak numbers were found to vary greatly between different experimental attempts for G4 ChIP-seq and do not represent a statistically robust method for assessing G4 ChIP signal.

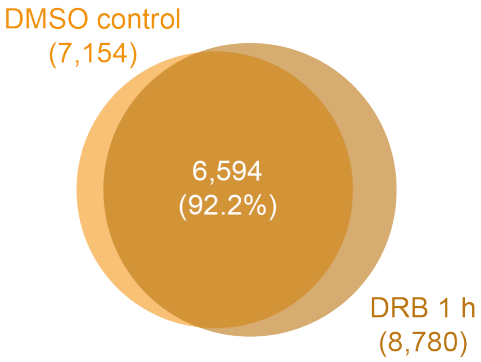

Figure 2.17: Overlap of G4 ChIP peaks between DRB-treated and DMSO-treated K562 cells.
Venn diagram showing overlap of consensus G4 ChIP-seq peaks in DMSO-treated and DRB-treated (100 µM, 1hour) K562 cells.

The signal differences in G4 ChIP-seq between DRB-treated and DMSO-treated cells were next statistically assessed. Only regions that showed 2-fold or more signal compared to input (i.e. with significant G4 ChIP signal enrichment) were selected as regions of interest (ROI). I first assessed the change of genome-wide G4 ChIP-seq signals. Within the 5,551 genome-wide G4 sites investigated, the majority of G4 sites (>95%) did not show statistically significant changes ($p < 0.01$) in the G4 ChIP signal (Figure 2.18A). Only 227 promoter G4 sites exhibited significantly decreased signal ($p < 0.01$) in response to DRB inhibition, and these regions were not considered for further analysis. I then further focused the analysis on G4s within gene promoters (4,023 sites). In a manner similar to genome-wide G4 response, promoter G4s in DRB- versus DMSO-treated cells also did not show statistically significant changes ($p < 0.01$) in the G4 ChIP signal at the majority of sites (Fig. 2.18B). These findings suggest that G4s may fold prior to productive elongation.

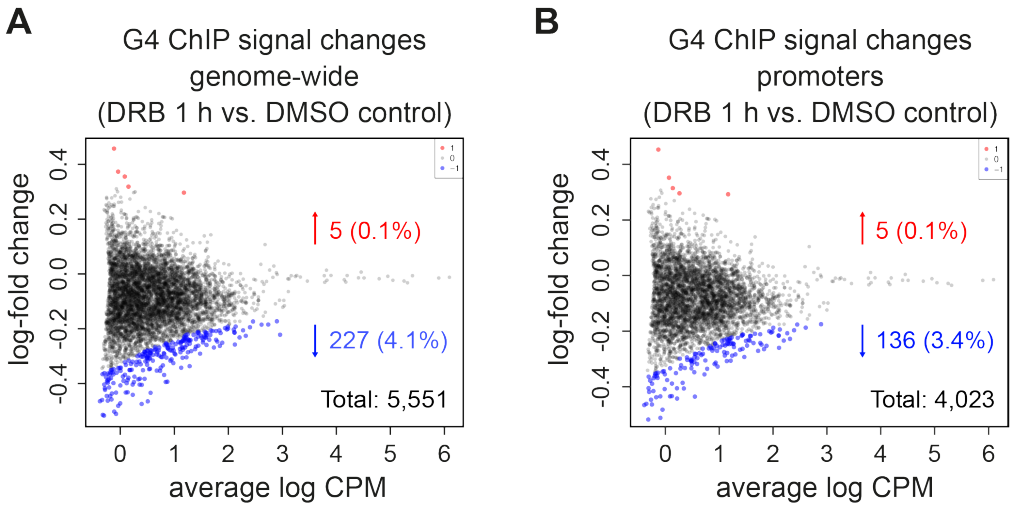

Figure 2.18: Majority of the sites show unchanged G4 ChIP-seq signal after DRB treatment.
(A) Bland Altman (MA) plot showing the fold change in G4 ChIP-seq signal at genome-wide regions of interest (ROI) between DRB-treated versus DMSO-treated K562 cells. G4 peaks showing statistically significantly ($p < 0.01$) increase and decrease in G4 ChIP-seq signals are highlighted in red and blue respectively; black dots indicate regions that do not change. CPM, read counts per million. **(B)** MA plot similar to panel A illustrating the fold change in G4 ChIP-seq signal at promoter G4s (TSS ± 500 bp; $p < 0.01$).

To assess G4 formation with an alternative methodology, BG4 immunofluorescence (IF) staining (Biffi et al., 2013b) was performed in parallel with G4 ChIP-seq (Figure 2.19). As a suspension cell line, K562 are not suited to IF staining. I therefore used U2OS cells, human bone osteosarcoma epithelial cells that have been widely used in BG4 IF staining. BG4 IF employs a three-layer signal amplification system, where BG4 signal in fixed permeabilised cells is bound by a secondary antibody against the FLAG tag, which is then further recognised by a tertiary fluorochrome-labelled antibody. This amplified signal was visualised and captured by confocal laser-scanning microscopy. Cells treated with DNase or performed in the absence of BG4 provided negative controls. Cells were also treated with RNase A to remove signal from RNA G4s. BG4 signal was seen as distinct foci primarily within nuclei, as has been reported previously (Biffi et al., 2013b). BG4 foci and signal intensity in each single cell nucleus were quantified using Icy (De Chaumont et al., 2012) and ImageJ respectively, and the median value with standard deviation (SD) was then calculated from the union of cell nuclei in each condition (Figure 2.19B). Similar to the number of BG4 foci reported in previous study (Biffi et al., 2013b), an average of 40 BG4 foci were identified in both conditions, i.e. DRB- and DMSO-treated U2OS cells, where no significant change was observed upon DRB treatment. In addition to this, BG4 signal intensity also did not change following DRB treatment (Figure 2.19C). These results further support our findings with G4 ChIP-seq, indicating that G4s are largely unaffected by DRB treatment.

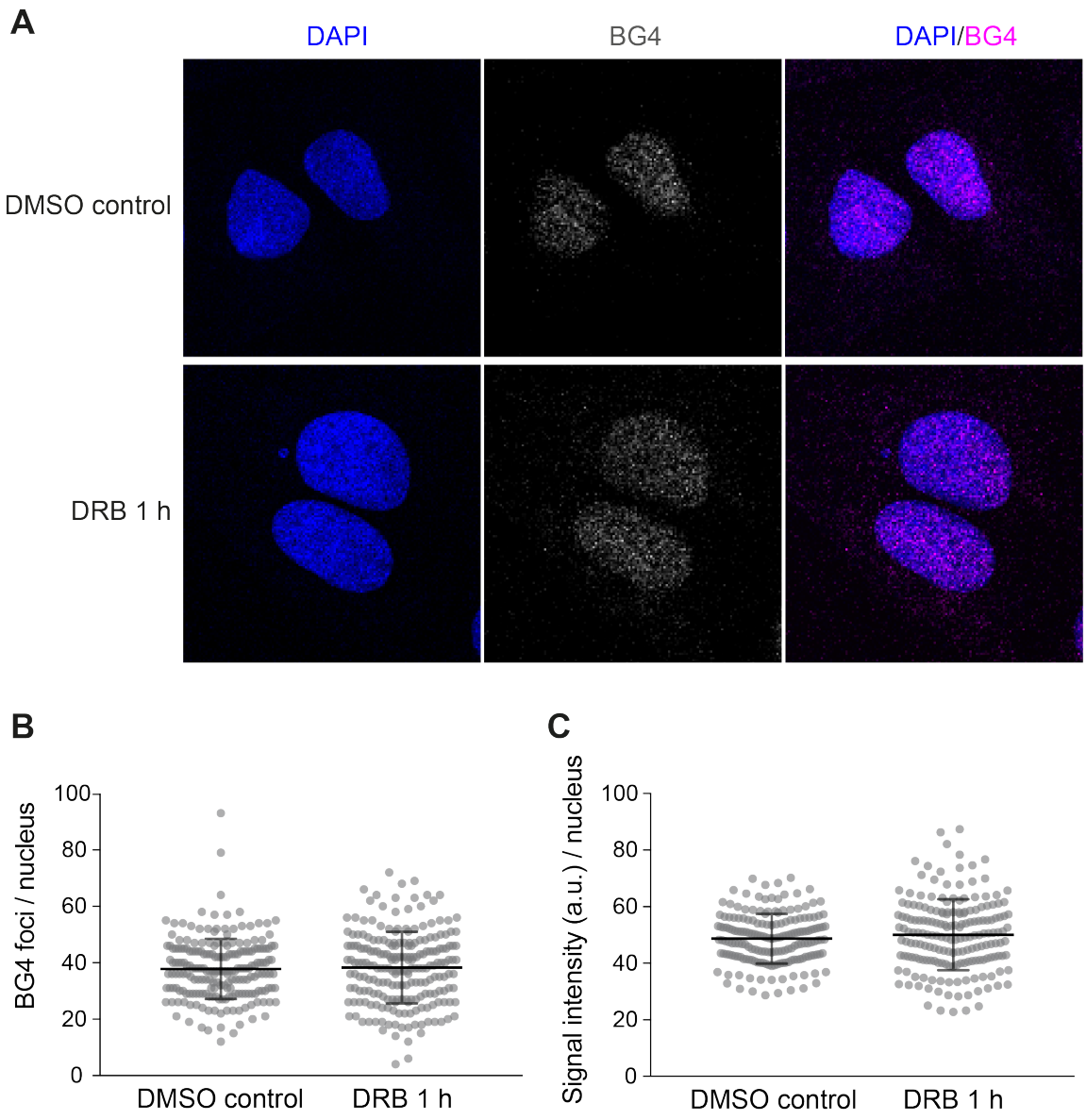

Figure 2.19: BG4 immunofluorescence signal is not altered following DRB treatment.
U2OS cells were paraformaldehyde fixed, RNase A treated and stained with BG4 antibody. Confocal images were captured using a Leica SP5 microscope. **(A)** BG4 staining (red) in the nuclei stained by DAPI (blue) of DRB-treated or DMSO-treated U2OS cells. **(B)** Graph quantifying the number of BG4 foci identified in individual cells. **(C)** Graph showing the average BG4 signal intensity in individual cells (a.u. = arbitrary unit). In panel B and C, each dot represents a single cellular nucleus. 150-200 nuclei were counted per condition and the Mann-Whitney test was performed across the mean values. Error bars represent the s.d. (n = 1).

2.3.6 Triptolide treatment inhibits transcription initiation and reduces cellular Pol II levels

Since inhibition of transcriptional elongation with DRB did not deplete promoter G4s, I next tested whether the inhibition of transcription initiation would impact the formation of G4 structures at promoters. For this, triptolide (TPL) was used to inhibit transcription initiation (Titov et al., 2011). During transcription initiation, the ATP-dependent activity of XPB is required to unwind the DNA double helix to allow RNA Pol II engage into active transcription. TPL covalently binds to XPB and inhibits its ATPase activity, preventing transcription initiation and ultimately inducing the degradation of RNA Pol II (Wang et al., 2011).

Previous reports have described a dose-dependent inhibition of transcription initiation by TPL (Chen et al., 2015a, 2015b; Erickson et al., 2018; Krebs et al., 2017). I therefore chose several TPL doses and times commonly used in literatures (125 nM for 1 hour, 10 μM for 30 minutes and 10 μM for 2 hours). The inhibition efficiency in K562 cells was determined by immunoblotting. TPL treatment resulted in a marked decrease in phosphorylation of serine 5 (Pol II Ser5-P) and serine 2 of RNA Pol II CTD (Pol II Ser2-P) (Figure 2.20A). Since Pol II Ser5-P marks active transcription initiation at TSSs (Eick and Geyer, 2013; Ni et al., 2008), the reduction of Pol II Ser5-P thus indicates the successful inhibition of transcription initiation. As reported previously (Wang et al., 2011), TPL treatment also caused a reduction in total RNA Pol II levels. To specifically compare between the different treatment regimes, protein bands in Figure 2.20A were quantified using densitometry (ImageJ) and the density of each band was normalised by loading control ß-actin (Figure 2.20B). The normalised protein levels were then plotted with levels in DMSO-treated K562 cells represented as 100%. 1-hour treatment with 125 nM TPL caused a marginal decrease in total (~15%) and phospho (~30%) RNA Pol II levels. By contrast, 30-min treatment with the higher dose of TPL (10 μM) led to a dramatic decrease (70%) in both. The 10 μM TPL treatment for 2 hours showed the strongest effect on inhibiting transcription in K562 cells. An over-90% decrease in RNA Pol II Ser5-P and RNA Pol II Ser2-P was detected, signifying that the majority of transcription initiation as well as transcription elongation had been halted. Over 80% reduction in total RNA Pol II protein levels were also observed, indicating the degradation of RNA Pol II in cells. Therefore, as reported, my experiments showed that TPL was found to inhibit transcription initiation and cause a reduction of RNA Pol II in a dose-dependent manner.

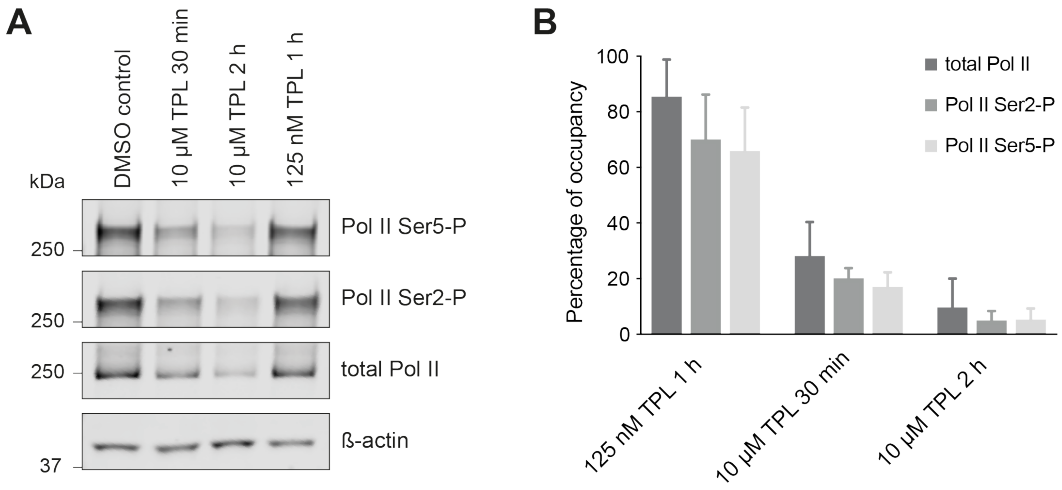

Figure 2.20: Triptolide (TPL) inhibits transcription initiation by RNA Pol II.
(A) Western blotting with antibodies against phosphorylated Pol II Ser2 (Pol II Ser2-P), phosphorylated Pol II Ser5 (Pol II Ser5-P), total Pol II and ß-actin in whole cell lysates from untreated K562 cells and TPL treated cells. ß-actin served as a loading control. Protein standards were used as molecular weight markers to determine protein size. **(B)** Quantification of protein levels detected in (A). The result was normalised by expression level of ß-actin and shown in occupancy proportion compared to untreated samples. Error bars represent the s.d. (n = 3).

2.3.7 Inhibition of transcriptional initiation does not lead to the loss of promoter G4 landscape.

I next explored how the endogenous G4 landscape, especially within gene promoters, is affected by the inhibition of transcription initiation. For this, I treated cells for 2 h with 10 µM TPL to achieve maximal transcriptional inhibition. G4 ChIP-seq from more than two independent biological replicates identified 6,173 and 10,631 consensus G4 peaks (i.e. G4 ChIP-seq peaks present in at least two biological replicates) in DMSO-treated and TPL-treated K562 cells respectively (Figure 2.21). The majority of the consensus G4 peaks (5,972/6,173; 96.7%) identified in DMSO-treated cells were also found in TPL-treated K562 cells. Although an increase of consensus peak number was observed after TPL treatment, for reasons explained in section 2.3.5, the robustness of the increase requires further validation.

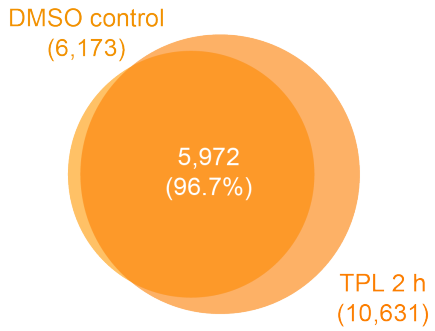

Figure 2.21: Overlap of G4 ChIP peaks between TPL-treated and DMSO-treated K562 cells.
Venn diagram showing positional overlap of consensus G4 ChIP-seq peaks in DMSO-treated and TPL-treated (10 μM, 2 hours) K562 cells. The majority of G4 peaks in DMSO-treated cells are found in TPL-treated cells.

The G4 ChIP signals from DMSO- and TPL-treated cells were quantified and compared. Using the criteria described in section 2.3.5, a total of 5,888 G4 sites were selected as genome-wide ROI for statistical comparison. Similar to DRB inhibition, TPL treatment only caused a statistically significant decrease in G4 ChIP signal at a very small number of sites ($p < 0.01$; 21/5888; Figure 2.22A). In contrast with DRB treatment, a significant increase in signal was observed at several sites ($p < 0.01$; 565/5,888, 9.6%) following TPL treatment.

Specifically, focusing the analysis on promoter G4s suggested a very similar trend to genome-wide G4s. While no reduction in G4 ChIP signal was observed following TPL treatment, ~10% of promoter G4s showed significantly increased G4 formation after the inhibition of transcription initiation with TPL ($p < 0.01$; 499/4,268, 10.5% up; Figure 2.22B). Since TPL inhibits the ATPase activity of Pol II-associated DNA helicases XPB and XPD, which function to resolve DNA G4s (Gray et al., 2014; Titov et al., 2011), my observation of increased promoter G4 formation following TPL treatment is likely to arise from the absence of such helicase activity. Importantly, chemical inhibition of transcription elongation or initiation does not cause reduced G4 folding at promoters and strongly suggests that active transcription does not contribute to G4 formation.

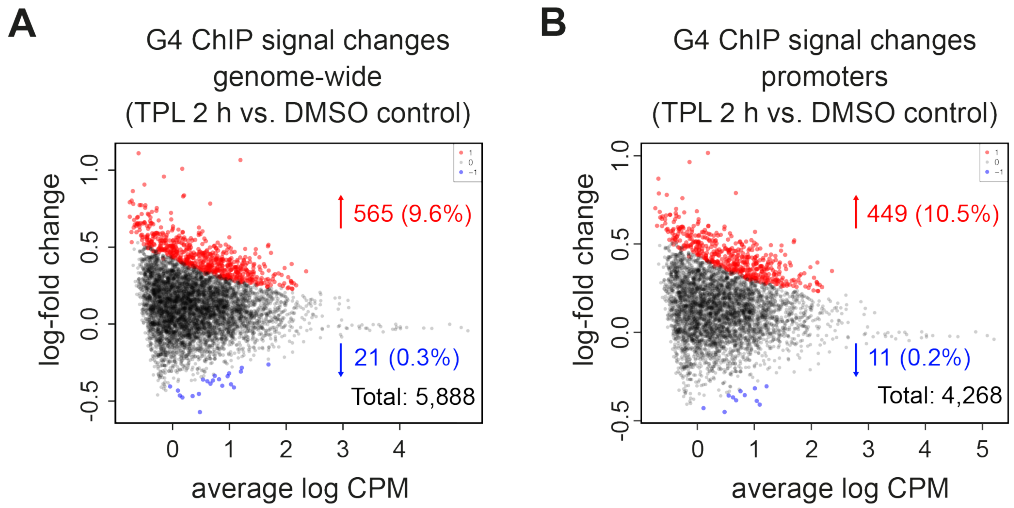

Figure 2.22: TPL treatment does not lead to decreases in G4 ChIP-seq signals.
(A) MA plot showing the fold change in G4 ChIP-seq signal at genomewide regions of interest (ROI) between TPL-treated versus DMSO-treated K562 cells. G4 peaks showing statistically significantly ($p < 0.01$) increase and decrease in G4 ChIP-seq signals are highlighted in red and blue respectively; black dots indicate regions that do not change. CPM, read counts per million. **(B)** MA plot similar to panel A illustrating the fold change in G4 ChIP-seq signal at promoter G4s (TSS ± 500 bp; $p < 0.01$).

As before, BG4 IF staining of DMSO- or TPL-treated U2OS cells was performed in parallel with G4 ChIP-seq. BG4 foci and signal intensity were quantified (Figure 2.23). The numbers of BG4 foci identified in TPL-treated U2OS cells remained the same as in DMSO-treated cells (40 foci/nucleus; Figure 2.23B). However, nuclei in TPL-treated cells showed a significantly higher BG4 signal intensity compared to DMSO-treated nuclei (DMSO: 28 versus TPL: 39; Figure 2.23C). This may be because BG4 IF lacks the spatial resolution to identify additional G4s as separate foci. Importantly, BG4 IF did not show any decrease in number of foci or signal, which is in agreement with the findings from G4 ChIP-seq and supports my contention that inhibition of transcriptional initiation does not deplete cellular G4s.

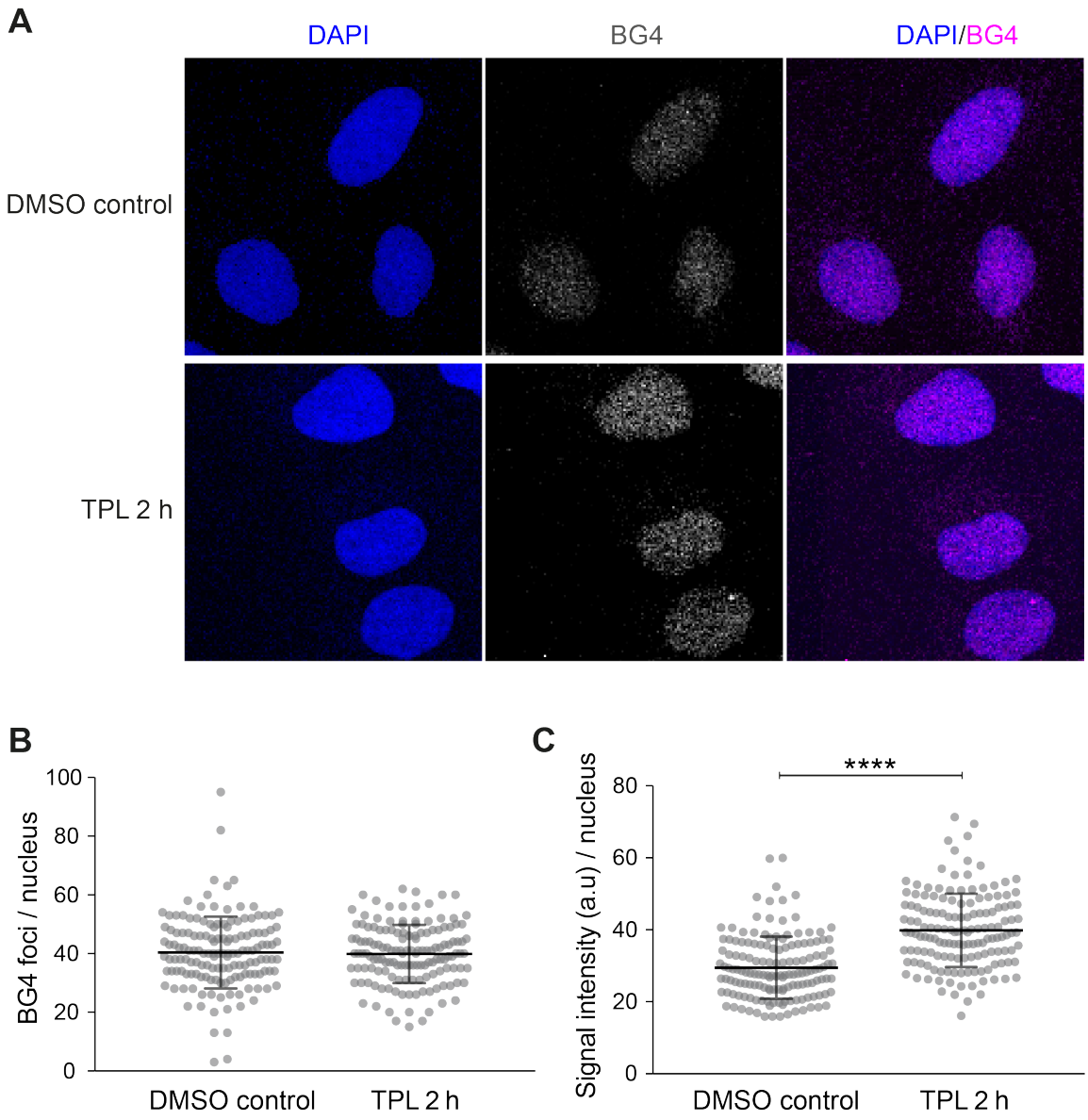

Figure 2.23: BG4 IF staining did not detect changes after TPL treatment.
U2OS cells were paraformaldehyde fixed, RNase A treated and stained with BG4 antibody. Confocal images were captured using a Leica SP5 microscope. **(A)** BG4 staining (red) in the nuclei stained by DAPI (blue) of TPL-treated or DMSO-treated U2OS cells. **(B)** Graph quantifying the number of BG4 foci identified in individual cells. **(C)** Graph showing the average BG4 signal intensity in individual cells (a.u. = arbitrary unit). ****: $p < 0.0001$. In panel B and C, each dot represents a single cellular nucleus. 150-200 nuclei were counted per condition and the Mann-Whitney test was performed across the mean values. Error bars represent the s.d. (n = 1).

2.4 Summary

It is established that G-rich sequences have the potential to fold into G-quadruplex structures in the human genome, which are enriched in functional regions such as gene promoters and 5'-UTRs (Chambers et al., 2015; Huppert and Balasubramanian, 2005, 2007; Kudlicki, 2016; Varizhuk et al., 2017). Endogenous G4 structures predominantly locate in nucleosome-depleted regions within promoters and mark sites of elevated transcription. However, the causal association between G4 formation and active transcription in cells has remained unknown. In this chapter, I have systematically studied the relationship between active transcription and promoter G4 formation in K562 cells to address whether the folding of G4s is driven by transcriptional activity.

I first purified and validated the BG4 antibody, which is employed in G4 ChIP-seq and BG4 IF staining throughout the study. Although BG4 preparation contained small amounts of protein contaminants from *E. coli*, these contaminants lack the 3 x FLAG tag that is used for subsequent applications and therefore seemed to have little impact on the binding affinity and specificity of BG4. The binding affinity determined by ELISA showed that BG4 bound to *in vitro* G4 structures with low nanomolar affinity, which is comparable to the affinity previously described (Biffi et al., 2013b). Furthermore, the G4 maps obtained with this BG4 batch showed comparable quality to previously published G4 ChIP-seq datasets, indicating that the contaminants do not interfere with this methodology.

In K562 cells, the majority of consensus G4s identified with G4 ChIP-seq were located in NDRs and were found highly enriched within gene promoters (defined as TSS ± 500 bp). These promoter G4s marked genes with significantly higher transcriptional output than promoters lacking a G4 structure. These results extend our original observations from HaCaT cells and primary keratinocytes (Hänsel-Hertsch et al., 2016), and establish the general principle that endogenous G4s are primarily found within open chromatin at promoters. Beyond that, G4-containing promoters also displayed greater RNA Pol II occupancy compared to their non-G4 counterparts in K562 cells, suggesting a correlation between the presence of promoter G4s with elevated RNA Pol II occupancy.

Transcription by RNA Pol II can result in increased ssDNA formation, either via torsional stress or by the formation of RNA/DNA hybrids (R-loops) (see section 1.6.3). Since formation of ssDNA can promote G4 folding (Duquette et al., 2004; Kouzine et al., 2004), it has been hypothesised that elevated transcription results in increased G4 formation at promoters. I directly tested this hypothesis by chemically inhibiting different stages of gene transcription and assessing G4 formation, especially at promoters by G4 ChIP-seq. If active transcription were to augment promoter G4 formation, a decrease in G4 ChIP signal would be expected following transcriptional inhibition. However surprisingly, promoter G4s in particular did not show a statistically significant decrease ($p < 0.01$) in the G4 ChIP signal upon inhibition of transcription elongation by DRB or upon inhibition of transcription initiation by TPL. Notably, both genome-wide G4s and promoter G4s showed a similar response to transcriptional inhibition, which indicates that the formation of G4s does not depend on active transcription regardless of the genomic location of G4 structures. A recent study has shown that DRB treatment abolishes ssDNA formation within the gene body, whereas TPL reduces ssDNA within the entire coding region (Wu et al., 2020). My findings that promoter G4s do not change upon DRB or TPL inhibition suggests that promoter G4 formation is independent from transcription-induced ssDNA.

Moreover, transcription inhibition by TPL partially led to increased G4 formation at promoters. TPL functions by inhibiting Pol II associated helicases XPB/XPD (Titov et al., 2011). XPD have been shown to resolve G4 structures *in vitro* and are found associated with G4 sites in chromatin (Gray et al., 2014; Hänsel-Hertsch et al., 2016). The inhibition of these helicases would therefore be expected to result in diminished resolution of secondary structures at promoters, including G4s, and explains the increase in G4 signal observed upon TPL treatment. Recently, Roychoudhury et al. mapped G4s in human adenocarcinoma cells using the monoclonal IgG form BG4 and reported that endogenous G4s are not reduced upon TPL treatment (Roychoudhury et al., 2020). However, they did not detect an increase in G4 ChIP signal that I observed in K562 cells, which may be due to differences in the cell line used or the ability to detect changes with the IgG version of BG4. This may also be due to lower concentration of TPL used by Roychoudhury et al. (500 nM) as opposed to 10 µM used in our study.

I further performed experiments supporting the findings of G4 ChIP-seq using BG4 IF staining. This method has been previously used to detect differences in G4 formation following treatment with G4 stabilising ligands or in cancer versus non-cancerous cells (Biffi et al., 2013b, 2014). Unlike G4 ChIP-seq that is able to detect G4 formation at specific sites, BG4 IF staining only provides a partial view of G4 foci that are detectable by IF. No significant decrease in BG4 signal was observed following transcriptional inhibition, which is consistent with results seen by G4 ChIP-seq. BG4 IF detects an average of 40 foci per nucleus, whereas G4 ChIP suggests the formation of thousands of G4s within the human genome. It is therefore plausible that each BG4 focus represents a cluster of closely-located endogenous G4s. Even though G4 IF has previously been used to detect large differences in G4 formation, for example upon treatment with G4 stabilising ligands (Biffi et al., 2013b), the ability of BG4 IF staining in detecting small changes in G4 population may be limited. It is interesting that TPL treatment elicits as an increase of BG4 signal intensity, which confirms the increase in G4 signal in TPL treated cells that was seen by G4 ChIP-seq. This may also support the idea that the G4 foci seen by IF mark clusters of individual G4s.

In summary, the G4 landscape in K562 cells described in this chapter corroborates previous observations in HaCaT cells and primary keratinocytes (Hänsel-Hertsch et al., 2016), which establishes G4s as a general feature of elevated gene expression across different cell types. This chapter also described that endogenous promoter G4s mark the elevated transcription activity and RNA Pol II occupancy, and that the folding of promoter G4s is not reliant on transcriptional activity. These findings contest earlier suggestions that the formation of promoter G4s is promoted by the generation of single stranded DNA during active transcription (Kouzine et al., 2004). Meanwhile, the findings in this chapter also raise the question of what drives G4 formation in cells. A previous study has suggested a correlation between chromatin accessibility and G4 formation (Hänsel-Hertsch et al., 2016). In that study, treatment of cells with inhibitors of histone acetylation caused some de-compaction of chromatin and some increases in G4 signal, suggesting that formation of G4s may be enhanced by increased chromatin accessibility. It is possible that chromatin context is more important for G4 formation than the active transcription. This possibility will be investigated and discussed in Chapter 3.

Chapter 3

The interplay between promoter G4s and chromatin structure

3.1 Background

Transcription is a strictly modulated cellular process, which is tightly linked with the dynamic regulation of chromatin remodelers, modification of histone tails, binding of transcription factors and activity of RNA polymerases (see section 1.5). While G4 structures are suggested to be involved in active transcription, growing evidence also indicates the interplay between G4s and chromatin structures. Early evidence from computational approaches predicted that regulatory G4s, such as those within promoters, are located outside of nucleosome-bound regions within the human genome (Wong and Huppert, 2009). In agreement with this, most endogenous G4s identified by G4 ChIP-seq overlapped with accessible open chromatin regions (Hänsel-Hertsch et al., 2016). Moreover, genome-wide potassium permanganate-dependent nuclease footprinting has also revealed that the formation of non-B DNA structures, including G4s, can in turn influence the positioning and occupancy of nucleosomes within chromatin (Kouzine et al., 2017).

Various studies have suggested that G4s interact with proteins that function to define chromatin architecture. For example, ATRX, a SWI/SNF-like chromatin remodeler, was found to bind a G4 from *NME* tandem repeat *in vitro* (Law et al., 2010). Mutational inactivation of *ATRX* induces the formation of G4 structures in glioma cells, and loss of ATRX sensitises cells to G4 ligands, including CX-3543 and PhenDC3 (Wang et al., 2019; Zyner et al., 2019). Moreover, the non-histone chromatin architectural protein HMGB1 has also been demonstrated to bind and stabilise the *KRAS* promoter G4 (Amato et al., 2018). Depletion of chromatin remodelers, such as SMARCA4, SMARCB1 and SMARCE1, was also found to sensitise cells to killing with G4 stabilising ligands, suggesting a functional interplay between chromatin architecture and G4s (Zyner et al., 2019). Further evidence also suggests an association between G4s and histone modifications. For instance, histone-lysine methyltransferase KMT5C, which methylates lysine 20 of H4 (H4K20) and causes chromatin condensation, has been reported to form a ternary complex with TERRA binding protein

TLS/FUS and bind telomeric G4 DNA (Takahama et al., 2013). Cellular G4s identified by IF staining using G4 antibodies, BG4 and IH6, has been shown to colocalise with trimethylation of H3 at lysine 4 (H3K4me3), a histone modification marks accessible chromatin and active transcription (Hänsel-Hertsch et al., 2016). Moreover, H4K20-specific histone demethylase PHF8 has been found enriched at G4 sites identified by G4 ChIP-seq (Hou et al., 2019).

In addition to these interactions, G4s have also been implicated in facilitating three-dimensional chromatin organisation. Topologically associated domains (TADs) play fundamental roles in regulating high-order chromatin structures (Ke et al., 2017). The precise formation of TAD boundaries, which is critical for gene regulation, is modulated by the binding of CCCTC-binding factor (CTCF) and cohesion proteins (Hong and Kim, 2017; Rao et al., 2014). Genomic G4 motifs, which conform to the canonical G4 sequence, were computationally determined to be located downstream of forward CTCF motifs and upstream of reverse CTCF motifs, whereas cohesion protein binding sites were found to be located in close proximity to G4s (Hou et al., 2019). This indicates a potential role of G4s in loop extrusion and preventing cohesin flow-through. Furthermore, it has also been suggested that two half G-quadruplexes (i.e. sequences comprising two runs of Gs) found within promoters and enhancers can form intermolecular G4s, which facilitates the promoter-enhancer interactions (Hegyi, 2015).

My observations in Chapter 2 indicate that the formation of promoter G4s as defined by G4 ChIP-seq is not promoted by active transcription, which raises the question as to what determines where G4s, specifically promoter G4s, form within cells. As determined in Chapter 2, G4s primarily form within accessible chromatin. It has also been shown that cells treated with entinostat, a small molecule that induces the acetylation of histone and cause de-compaction of chromatin, displayed elevated transcriptional output identified by RNA-seq and some increases in G4s identified by G4 ChIP-seq (Hänsel-Hertsch et al., 2016). This suggests that formation of G4s is correlated with increased chromatin accessibility and/or transcriptional activity. Since I have shown that promoter G4 formation does not require transcriptional activity, I therefore hypothesise that permissive chromatin is a determinant of promoter G4 formation. To test this, I will ask in this chapter whether chromatin compaction has any impact on the cellular G4 landscape.

Histone acetylation is critical for modulating chromatin structure, and is regulated by two opposing classes of histone modifying enzymes, histone acetyltransferases (HATs) and histone deacetylases (HDACs) (Bannister and Kouzarides, 2011; Wang et al., 2009) (see section 1.5.1). Acetylation of nucleosomal histones can directly or indirectly result in alteration of chromatin structure (Davie and Moniwa, 2000). Ectopic overexpression of HDACs or inhibition of HAT activity can lead to reduced histone acetylation to impact chromatin condensation. Several studies have shown that cellular response to environmental stress stimuli can also trigger chromatin compaction. In particular, the cellular response to acute oxygen depletion is well-documented for inducing genome-wide reduction in chromatin accessibility (Batie et al., 2018). For example, acute hypoxia is accompanied by elevated methylation at lysine 9 of histone H3 (H3K9me3) (Chen et al., 2006) and reduced histone acetylation (Gao et al., 2016). H3K9me3 marks heterochromatin regions and recruits chromatin modifying proteins, such as heterochromatin proteins (Batie et al., 2018; Hyun et al., 2017; Ninova et al., 2019). Indeed, cells exposed to hypoxia showed increased expression of the heterochromatin protein HP1BP3 (Dutta et al., 2014). In addition, the study also found that hypoxic cells have reduced sensitivity to mononuclease digestion, suggesting chromatin compaction. Therefore, in this chapter, I investigate different approaches, in particular HAT inhibition and hypoxia, to rapidly cause chromatin compaction in cells, and evaluate their impact on the cellular G4 landscape using G4 ChIP-seq (Figure 3.1).

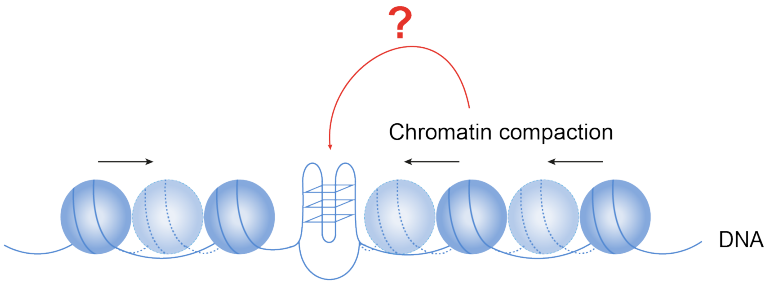

Figure 3.1: Exploring the interplay between promoter G4s and chromatin structure.
Graphical representation of experimental design examining the response of cellular G4s to reduced chromatin accessibility. Chromatin compaction is induced by HAT inhibition or hypoxia and the change of G4s is subsequently assessed by G4 ChIP-seq.

3.2 Objectives

The aim of this chapter is to explore the interplay between endogenous G4s and chromatin structure. My first objective is to establish a method for inducing rapid and reliable chromatin compaction in K562 cells. I will test HAT inhibitors and hypoxia treatment for this purpose. To minimise the influence of possible indirect downstream effects on promoter G4 folding, I choose acute short treatments. The second objective of this chapter was to assess the folding of promoter G4s in K562 cells following chromatin compaction. The third objective is to analyse the influence of reduced chromatin accessibility under conditions of chromatin compaction on RNA Pol II occupancy to further tease apart the relationship of G4s with chromatin and transcription.

3.3 Results

3.3.1 The effect of histone acetyltransferase inhibitors on histone acetylation

HAT inhibition causes reduced acetylation of histones, and consequently promotes chromatin condensation (Gregory et al., 2001). It has been suggested that many commonly used HAT inhibitors are thiol-reactive and/or protein aggregators (Dahlin et al., 2017), which may lead to non-specific downstream effects in cells. This emphasises the need to refrain from longer treatment times in order to limit off-target or downstream impact on cellular G4 landscape. To investigate maximal reduction in histone acetylation within a short time, I first treated K562 cells for short durations with commercially available HAT inhibitors. I tested compounds that target different HATs, including anacardic acid, curcumin and garcinol that are p300/CBP-associated HAT inhibitors (Balasubramanyam et al., 2003, 2004; Marcu et al., 2006), as well as CPTH2 and CPTH6 which are GCN5 inhibitors (Chimenti et al., 2009; Trisciuoglio et al., 2012). I first tested their impact on pan-histone acetylation using Western blotting with antibodies specific for acetylated forms of histone H3 and histone H4 (Figure 3.2). However, following treatments for 2 or 4 h, only marginal changes were observed for H3 and H4 acetylation. Importantly, 4 h treatment of K562 cells with 100 µM garcinol led to severe cell death, whereby enough lysate could not be obtained for Western blotting. I therefore further explored if a lower concentration of garcinol could deplete histone acetylation as discussed in section 3.3.2.

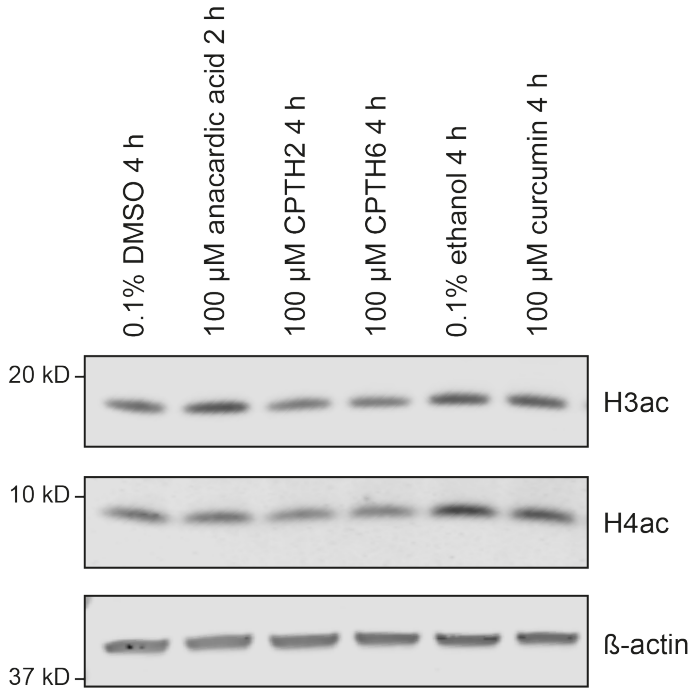

Figure 3.2: HAT inhibitors does not dramatically alter the cellular histone acetylation levels.
Western blotting for acetylated histone H3 (H3ac) and acetylated histone H4 (H4ac) from K562 cells treated with HAT inhibitors. ß-actin provides the loading control. The ethanol provides the control treatment for curcumin, whereas DMSO provides the control treatment for the remaining HAT inhibitors.

3.3.2 Garcinol-mediated histone deacetylation causes chromatin compaction

To evaluate how acetylation of histone H3 and H4 changes upon garcinol treatment, K562 cells were treated with different concentrations of garcinol (0.625 µM to 5 µM) for 4 h. Reducing the concentration of garcinol reduced the level of cell death that was previously observed with 100 µM garcinol and allowed me to obtain sufficient cell lysates for Western blotting. Garcinol treatment showed a dose-dependent decrease in histone H3 and H4 acetylation (Figure 3.3). Notably, 5 µM garcinol leads to a substantial reduction of H3ac and H4ac. This is in agreement with a previous study where treatment with 5 µM garcinol caused depletion of histone H3 acetylation in *Toxoplasma gondii* (Jeffers et al., 2016).

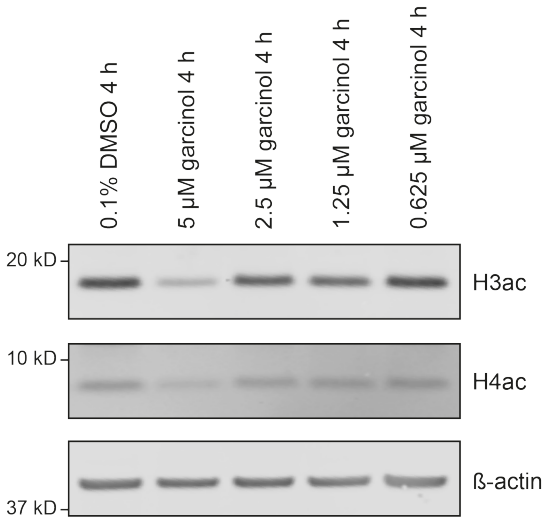

Figure 3.3: Garcinol treatment causes reduced histone H3 and H4 acetylation in cells.
Western blotting for acetylated histone H3 (H3ac) and acetylated histone H4 (H4ac) from K562 cells treated with garcinol at different concentrations. Cells were treated with garcinol for 4 h. ß-actin provides the loading control. DMSO provides the negative control treatment. The protein levels of H3ac and H4ac shows substantial reduction after the treatment with 5 µM garcinol for 4 hours.

To determine if garcinol-mediated histone deacetylation in K562 cells affects chromatin structure, the accessibility of chromatin was first examined by measuring the accessibility to digestion by Micrococcal Nuclease (MNase). Nuclei extracted from cells were treated with 0.5 units of MNase and reaction stopped at different times. Digestion by MNase requires DNA not to be bound to nucleosomes, so that the enzyme can cleave accessible DNA in chromatin more effectively than compacted chromatin that is bound by more nucleosomes. Digested DNA fragments were subsequently separated on an agarose gel, where a typical nucleosomal “ladder” pattern could be observed (Figure 3.4A). Longer digestion (45 min) causes the majority of genomic DNA to resolve as mono-nucleosomal DNA (1n). But at earlier digestion times, a small but clear difference could be observed in the amount of poly-nucleosomal DNA in garcinol treated cells when compared to DMSO. To directly visualise the difference, band intensities corresponding to mono- and poly-nucleosomes for the 10 min time-point were plotted as signal density (Figure 3.4B). Compared to DMSO condition, less mono-nucleosomes but more poly-nucleosomes were observed upon garcinol treatment, suggesting that given the same incubation time with MNase, garcinol-treated cells were less sensitive to MNase digestion compared to DMSO-treated cells. These results indicate that chromatin become more compacted upon garcinol treatment.

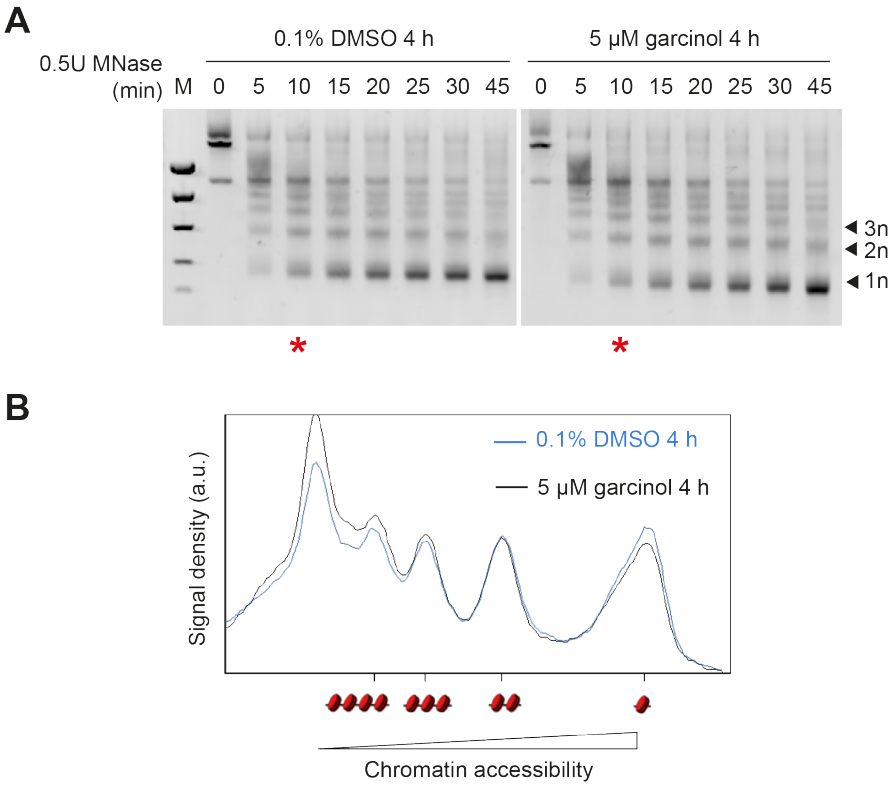

Figure 3.4: Micrococcal Nuclease (MNase) digestion revealed increased chromatin compaction in garcinol-treated K562 cells.

(A) Electrophoresis analysis showing the nucleosome distribution for DMSO- or garcinol-treated K562 cells digested by MNase for increasing digestion times. 1n = mono-nucleosomes, 2n = di-nucleosomes, 3n = tri-nucleosomes. Asterisks indicate samples quantified by the graph in **(B)**. A decrease in mono-nucleosomal fragments and an increase in poly-nucleosomal are seen in garcinol-treated cells (black) when compared to DMSO-treated control cells (blue).

To further examine chromatin accessibility upon garcinol treatment, ATAC-seq was performed. ATAC-seq can provide information on chromatin accessibility at specific genomic locations. The size of fragments incorporated into an ATAC-seq library can provide a genome-wide measure of chromatin accessibility. Fragments from ATAC-seq libraries were binned into different sizes and the fragment count for each bin plotted as shown in Figure 3.5. Mono-nucleosomal fragments are observed as a peak centered at 200 bp, di-nucleosomal fragments as peak centered at 400 bp. This periodic pattern is well-established in previous reports (Buenrostro et al., 2013; Yan et al., 2020). The increase in size of DNA fragments protected by nucleosomes is due to the spacing (+ ~38 bp) required by the transposition event (Adey et al., 2010). A clear reduction in mono-nucleosomal fragments and an increase in longer fragments was observed upon garcinol treatment, which indicates reduced accessibility to the hyperactive Tn5 transposase (Figure 3.5). This result is consistent with the findings from the MNase digestion assay and further supports that garcinol causes genome-wide chromatin compaction.

In addition to global fragment size analysis, to examine the chromatin accessibility at promoters upon garcinol treatment, changes in ATAC signal intensity at promoters was compared in DMSO- and garcinol-treated cells. In control conditions, gene promoters marked by G4 ChIP-seq peaks (i.e. $G4^+$) were found to demonstrate a higher accumulation of ATAC signal when compared to promoters without a G4 ChIP-seq peak (i.e. $G4^-$; Figure 3.6A). Since G4s primarily form in open chromatin regions (see section 2.3.2; Figure 2.12), these findings show that $G4^+$ promoters generally feature more accessible chromatin than their non-G4 counterparts. Upon treatment of cells with garcinol, a reduction in ATAC signal was observed at most promoters, but $G4^+$ gene promoters showed a greater reduction when compared to $G4^-$ counterparts (Figure 3.6B). This suggested that chromatin at G4 marked promoters was more compacted than at non-G4 promoter. Remarkably, of 6,856 $G4^+$ promoters, more than 20% showed a significant reduction in ATAC-seq signal intensity in garcinol-treated K562 cells ($p < 0.05$; 1,640/6,856 down; Figure 3.7). This indicates that within the short treatment period, garcinol is able to cause substantial chromatin compaction at promoter G4s.

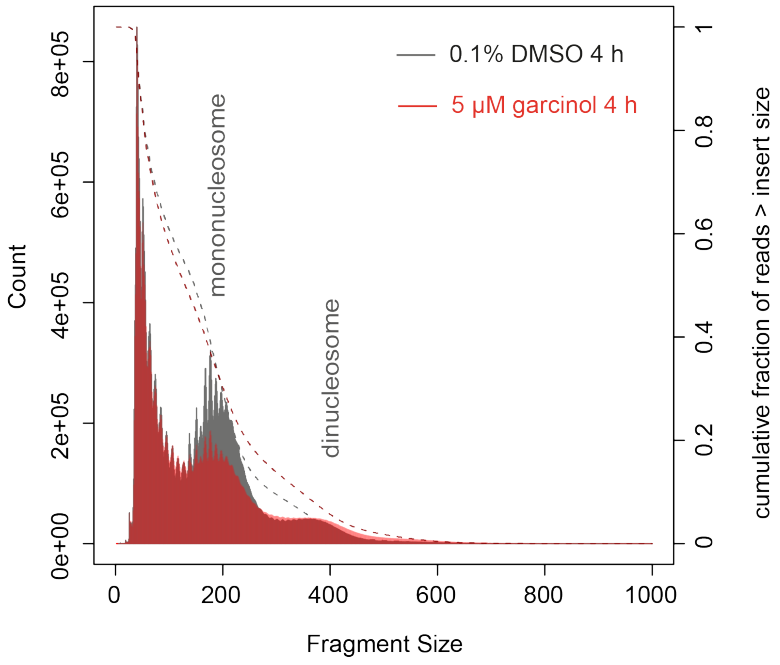

Figure 3.5: Garcinol-induced HAT inhibition increases fragment sizes of ATAC-seq libraries.
Read counts and size distribution for different ATAC-seq fragment sizes for DMSO-treated (grey) and garcinol-treated (red) chromatin. Mono-nucleosomal fragments are observed as a peak centered at 200 bp and di-nucleosomal fragments as peak centered at 400 bp. The reduction of mono-nucleosomal fragments and the increase in longer fragments observed upon garcinol treatment suggests that garcinol causes genome-wide chromatin compaction.

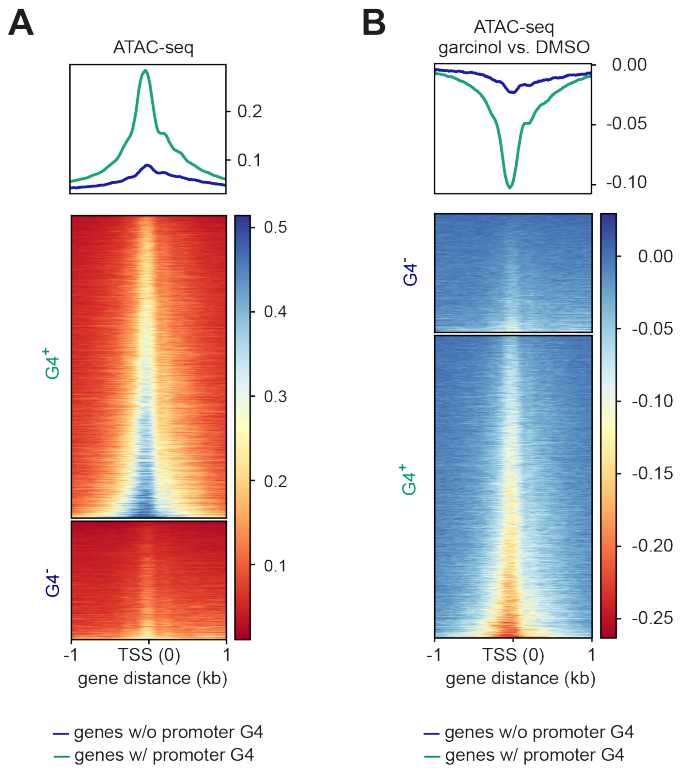

Figure 3.6: Promoter G4s show greater chromatin accessibility but more compaction upon garcinol treatment.

(A) Chromatin accessibility at promoters in DMSO treated control cells. Top panel, metagene plot of the normalised ATAC signal centered at the transcription start sties (TSS). Green, promoters marked with G4 ChIP-seq peaks ($G4^+$); blue, promoters that do not have G4 ChIP-seq peaks ($G4^-$). Bottom panel, normalised ATAC-seq signal plotted for all individual loci and represented by a heatmap. $G4^+$ promoters feature more accessible chromatin than their non-G4 counterparts in normoxia. **(B)** Signal differences in ATAC-seq signal for genes with (green) or without (blue) a promoter G4 under garcinol or DMSO. Top panel, metagene plot of the normalized ATAC signal centered at the TSS. Bottom panel, normalised ATAC-seq signal plotted for all individual loci and represented by a heatmap. Chromatin at $G4^+$ promoters is more compacted than at $G4^-$ promoters upon garcinol treatment.

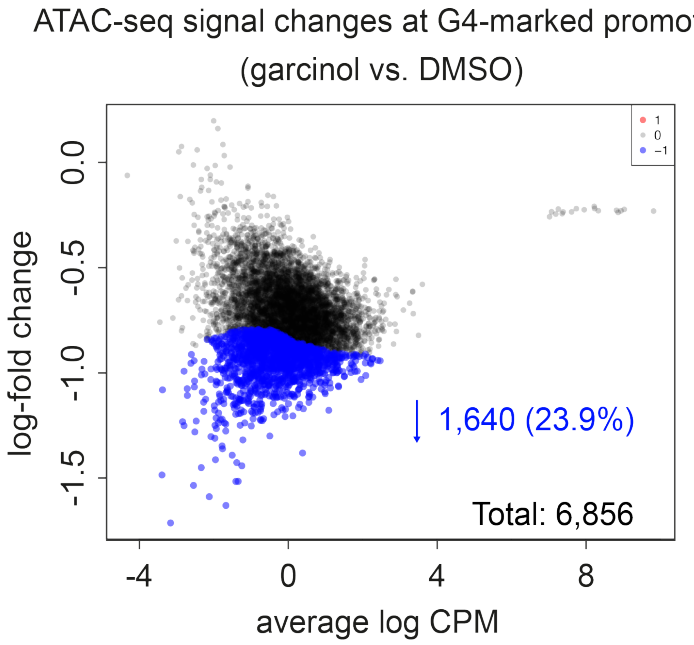

Figure 3.7: A fifth of G4 marked promoters show chromatin compaction after garcinol induction.
MA plot showing fold change in ATAC-seq signal following garcinol treatment (5 μM, 4 h) at G4-marked gene promoters. Blue, sites with significantly ($p < 0.05$) decreased signal following garcinol treatment. Black, sites with unchanged signal following garcinol treatment. CPM, read counts per million.

3.3.3 Garcinol treatment increases cellular G4 content

I next examined how G4 landscape alters in response to the induction of chromatin compaction by garcinol treatment. To evaluate the change in G4 ChIP signal at promoters, G4 ChIP-seq was performed on garcinol-treated cells. Consensus G4 peaks were first filtered to include only those that show 2-fold greater signal when compared to input. G4 ChIP signal at filtered promoters was then compared in DMSO versus garcinol treated cells. Garcinol treatment did not significantly decrease the G4 ChIP-signal ($p < 0.05$; 5/2,863 down, 0.2%; Figure 3.8). In contrast, several sites demonstrated a significant increase in G4 formation ($p < 0.05$; 275/2,863, 9.6%), suggesting that G4 formation can occur in the presence of garcinol.

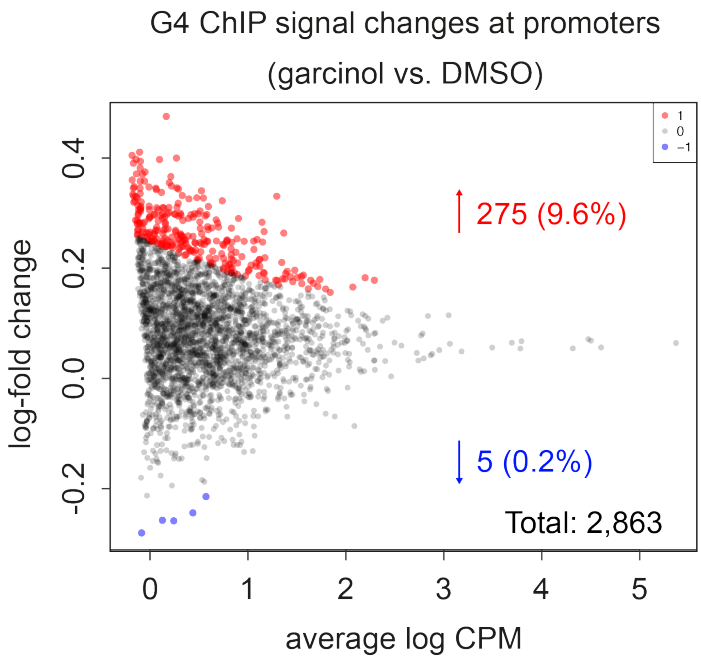

Figure 3.8: G4 ChIP-seq signal increases upon garcinol treatment.
MA plot showing the fold change in G4 ChIP-seq signal at promoters (TSS ± 500 bp) in garcinol-treated versus DMSO-treated K562 cells. Statistically significantly ($p < 0.05$) higher and lower signals are in red and blue respectively; black dots indicate regions not changing. CPM, read counts per million.

Garcinol treatments have been shown to induce oxidative stress in human erythrocytes (Fazio et al., 2015). A key enzymatic component of the base excision repair (BER) pathway, APE1, has been showed to drive G4 formation in response to DNA damage (Roychoudhury et al., 2020). It is thus plausible that the observed increase in promoter G4 formation is due to the induction of BER following garcinol treatment.

To examine whether the increased G4 formation on garcinol treatment could be caused by activation of BER, Western blotting was performed to assess APE1 activation upon garcinol treatment. APE1 becomes acetylated (acAPE1) on binding abasic sites (AP sites) in DNA that are caused by oxidative damage. Methoxyamine binds to the aldehyde in AP sites, which is recognised by APE1 during BER, and thus inhibits the binding and activation of APE1 (Hegde et al., 2012; Liu and Gerson, 2004). By contrast, hydrogen peroxide (H_2O_2) triggers oxidative DNA damage and promotes the endonuclease and redox activities of APE1 in cells (Pines et al., 2005; Tell et al., 2005). Treatment of K562 cells with H_2O_2 showed an increase in acAPE1

protein levels, whereas very low levels of acAPE1 was detected in methoxyamine-treated K562 cells as seen by Western blotting (Figure 3.9). Although the standard ß-actin loading control showed an upward band shift following H_2O_2 treatment, which is likely to be due to its polyubiquitination (Reeg et al., 2020), the total protein level of ß-actin remains similar. Therefore, the ß-actin protein is still used as the loading control. Garcinol-treated cells showed an increase in acAPE1 protein level when compared to DMSO and methoxyamine treatments. This indicates that garcinol treatment triggers significant APE1 activation and may explain the increase in G4s following garcinol treatment. The interaction between APE1 and chromatin compaction is unclear, thereby ruling out garcinol as a method for addressing my hypothesis.

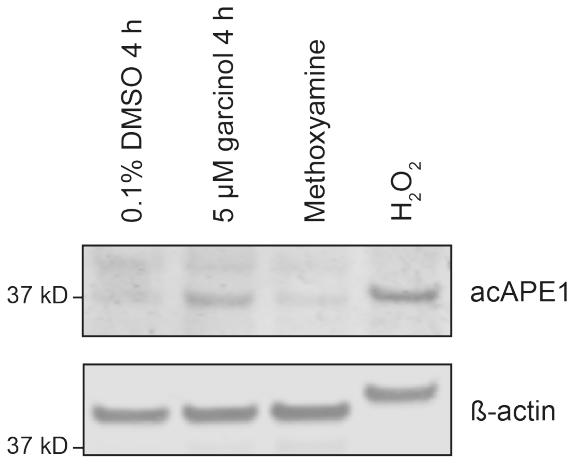

Figure 3.9: Garcinol treatment induces APE1 activation.
Western blotting for acetylated APE1 (acAPE1) in DMSO- or garcinol-treated K562 cells. The increase of acAPE1 protein levels suggests active APE1 binding and repair. Methoxyamine- and H_2O_2-treated cells represent negative and positive control treatments respectively. ß-actin is the loading control.

3.3.4 Acute Hypoxia causes global chromatin compaction

As an alternative to HAT inhibition, I investigated acute hypoxia as a method to trigger chromatin condensation (Chen et al., 2006; Dutta et al., 2014; Gao et al., 2016; Kirmes et al., 2015). To minimise possible downstream indirect effects on promoter G4s, K562 cells were exposed to hypoxia in a relatively short period of time (1% oxygen, 1 h; referred to as acute hypoxia) compared to normoxia (21% O_2). Reduced oxygen induces the upregulation of hypoxia inducible factor 1α (HIF1α) (Batie et al., 2018). I first confirmed that K562 cells exposed to acute hypoxia showed an increase in HIF1α protein level as seen by Western blotting. Figure 3.10A indicates successful induction of cellular hypoxic response in this cell line.

Next, to assess chromatin accessibility, MNase digestion was performed for cells grown under normoxic and hypoxic conditions. As before (see section 3.3.2), extracted nuclei were incubated with 0.5 unit MNase from 2 min to 20 min, and digested DNA was separated by agarose gel electrophoresis (Figure 3.10B). Reduced intensity of the mono-nucleosomal band was observed in hypoxic cells when compared to normoxic cells at every individual digestion time-point. The difference of band intensities corresponding to mono- and poly-nucleosomes for the 20 min time-point between normoxia and hypoxia were plotted as signal density (Figure 3.10C). Hypoxia was found to cause substantial reduction in mono-nucleosomal fragments and an increase in poly-nucleosomal fragments when compared to normoxia. This indicates that hypoxic nuclei had reduced sensitivity to MNase digestion due to chromatin compaction. In parallel, ATAC-seq was performed to examine chromatin accessibility upon hypoxia. Fragment size analysis of ATAC-seq data also showed an increase in overall fragment size following hypoxia when compared to normoxia, indicating reduced accessibility to the hyperactive Tn5 transposase (Figure 3.10D). The fragment size differences observed by ATAC-seq further corroborates the results from MNase digestion, and demonstrates genome-wide chromatin compaction following acute hypoxia.

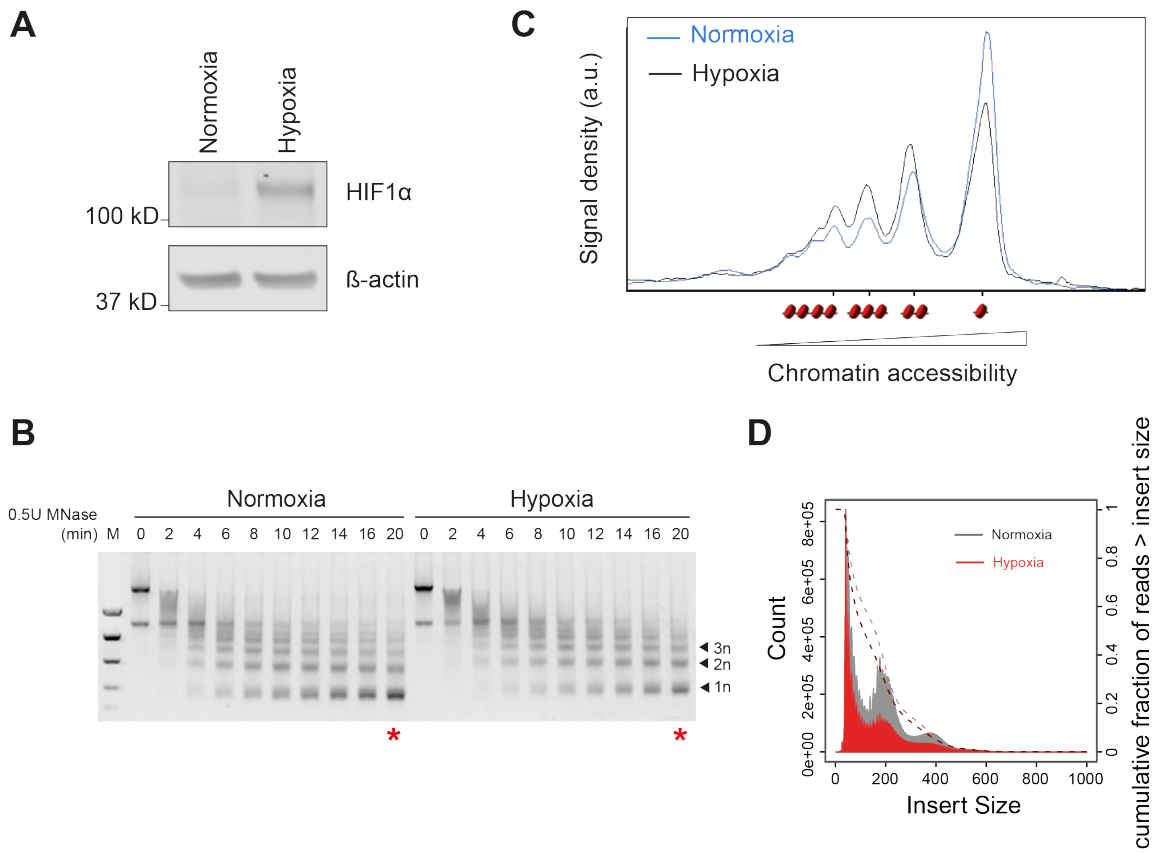

Figure 3.10: Oxygen depletion leads to HIF1α induction and chromatin compaction.
(A) Western blotting for HIF1α from K562 cells cultured in normoxia and hypoxia. Actin provides the loading control. The hypoxic response that is marked by HIF1α in K562 cells is successfully induced. **(B)** Agarose gel electrophoresis of DNA from normoxic or hypoxic K562 chromatin digested with 0.5-unit MNase for increasing time. 1n = mono-nucleosomes, 2n = di-nucleosomes, 3n = tri-nucleosomes. Asterisks indicate lanes scanned and quantified by the graph in **(C)**. A substantial reduction in mono-nucleosomal fragments and an increase in poly-nucleosomal are seen in hypoxia (black) when compared to normoxia (blue). **(D)** Read counts and size distribution for different ATAC-seq fragment sizes for chromatin under normoxia (grey) and hypoxia (red). These findings suggest that acute hypoxia treatment causes global chromatin compaction in K562 cells.

3.3.5 Hypoxia induces chromatin compaction at most promoter G4s

Having shown that acute hypoxia causes global chromatin compaction, I specifically investigated the chromatin status at G4s within promoters of active genes as defined by RNA Pol II occupancy (i.e. co-occurrence of G4 ChIP-seq peaks and RNA Pol II ChIP-seq peaks within TSS ± 500 bp). Similar to my earlier observations (see section 3.3.2; Figure 3.6), G4-marked active gene promoters (i.e. Pol II^+ $G4^+$) showed greater chromatin accessibility as seen by increased ATAC-seq signal intensity when compared to their non-G4 marked (Pol II^+ $G4^-$) active gene counterparts (Figure 3.11A). Upon hypoxia, a greater reduction in ATAC signal was observed at Pol II^+ $G4^+$ promoters than non-G4 counterparts (Pol II^+ $G4^-$) (Figure 3.11B). This suggests that although hypoxia induces global chromatin condensation, G4-marked active promoters are overall particularly sensitive. Remarkably, of the 9,217 Pol II^+ $G4^+$ promoter sites considered, over 85% of them showed a significant reduction in ATAC-seq signal intensity in hypoxic K562 cells ($p < 0.05$; 7,920/9,217 down, 85.9%; Figure 3.12).

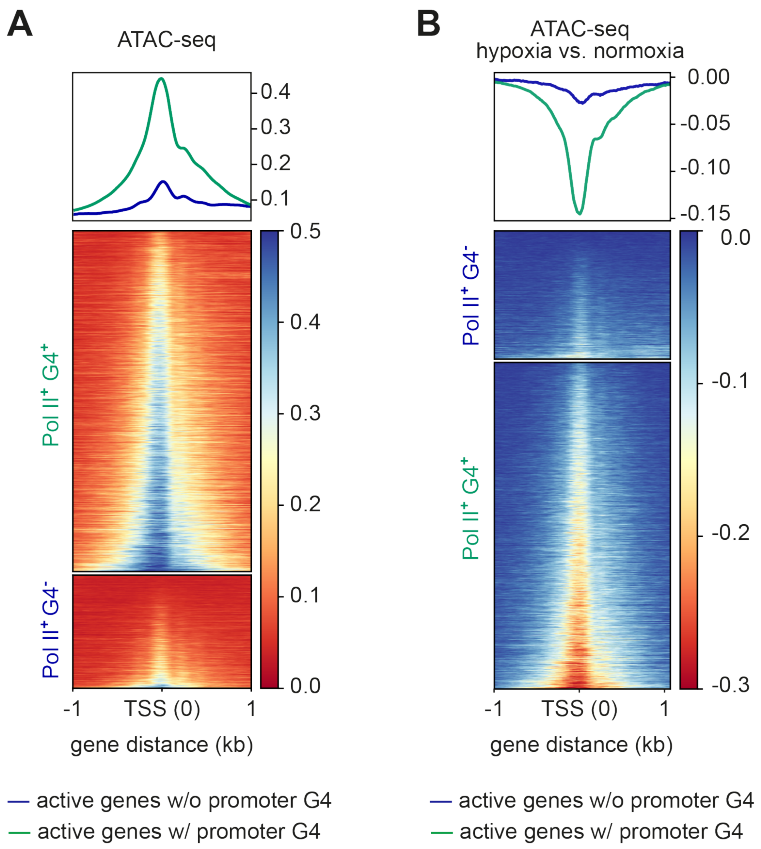

Figure 3.11: G4-marked active promoters show greater chromatin accessibility in normoxia than in hypoxia. **(A)** Chromatin accessibility at promoters in normoxia as measured by ATAC-seq. Top panel, metagene plot of the normalised ATAC signal centered at the TSS. Green, G4 marked active gene promoters (Pol II^+ $G4^+$); blue, non-G4-marked active gene promoters (Pol II^+ $G4^-$). Bottom panel, normalised ATAC-seq signal plotted for all individual loci and collectively represented as a heatmap. Pol II^+ $G4^+$ promoters feature more accessible chromatin than their non-G4 counterparts in normoxia. **(B)** Signal differences in ATAC-seq signal for active genes with (green) or without (blue) a promoter G4 in hypoxia versus normoxia. Top panel, metagene plot of the normalised ATAC signal centered at the TSS. Bottom panel, normalised ATAC-seq signal plotted for individual loci and collectively represented by a heatmap. Normoxia refers to cells cultured in 21% O_2 and hypoxia refers to cells exposed to 1% O_2 for 1 h. Chromatin at Pol II^+ $G4^+$ promoters becomes more compacted than at the Pol II^+ $G4^-$ counterparts upon hypoxia.

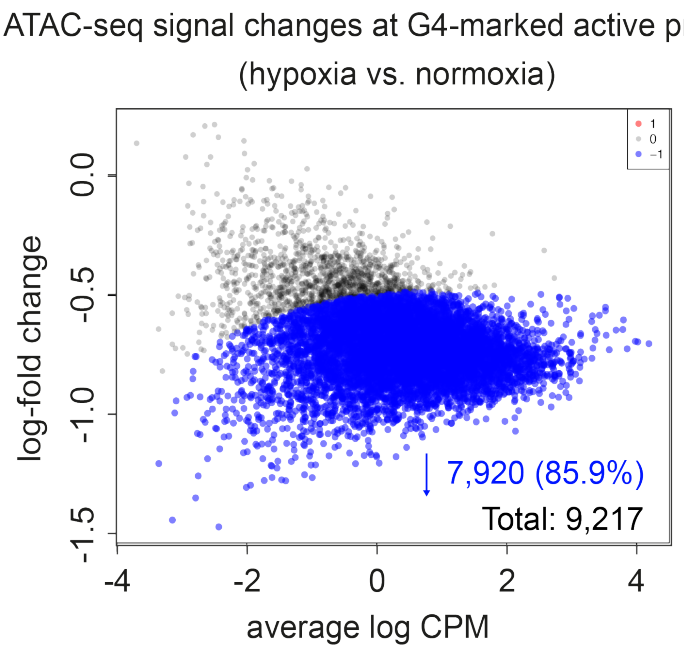

Figure 3.12: Hypoxia induces chromatin compaction at the majority of G4-marked active gene promoters. MA plot showing fold change in ATAC-seq signal following hypoxia at G4-marked active gene (Pol II^+ $G4^+$) promoters. Blue, sites with significantly ($p < 0.05$) decreased signal. Black, sites that are not changed in ATAC-seq signal intensity. CPM, read counts per million. Normoxia refers to cells cultured in 21% O_2 and hypoxia refers to cells exposed to 1% O_2 for 1 h. The majority of sites are seen to become more compacted in hypoxia relative to normoxia.

3.3.6 Promoter G4 formation is sensitive to chromatin compaction

To assess the impact of chromatin compaction via hypoxia on the formation of promoter G4s, G4 ChIP-seq was performed in normoxic and hypoxic chromatin. As described previously (see section 3.3.3 and section 3.3.5), consensus G4 peaks at active promoters (i.e. Pol II^+ $G4^+$ promoters) were filtered for those that show greater than 2-fold signal compared to the corresponding input. The G4 ChIP-seq signal at the filtered regions were compared between hypoxia versus normoxia. Strikingly, nearly 20% of these filtered promoter G4s showed a statistically significant reduction in signal intensity ($p < 0.05$; 1,105/5,558, 19.9%; Figure 3.13), indicating that the formation of many promoter G4s is sensitive to chromatin compaction. To validate these findings by an alternative methodology, nuclear G4 foci were visualised in U2OS cells, which show the most cellular G4 foci compared to other cell lines tested, using BG4 IF staining (Biffi et al., 2013b) (Figure 3.14). Hypoxia was found to cause a significant decrease in the number of BG4 foci per nucleus when compared to normoxia ($p <$

0.001; 100-150 nuclei from 1 biological replicate was compared; normoxia: 43 vs. hypoxia: 39; Figure 3.14B). Nuclei in hypoxic cells also showed a significantly lower BG4 signal intensity compared to normoxic nuclei ($p < 0.0001$; normoxia: 46 vs. hypoxia: 38; Figure 3.14C). Together, BG4 IF showed decreases in both BG4 foci number and signal intensity, which supports with the findings of G4 ChIP-seq and confirms that many endogenous G4s are sensitive to chromatin compaction induced by acute hypoxia.

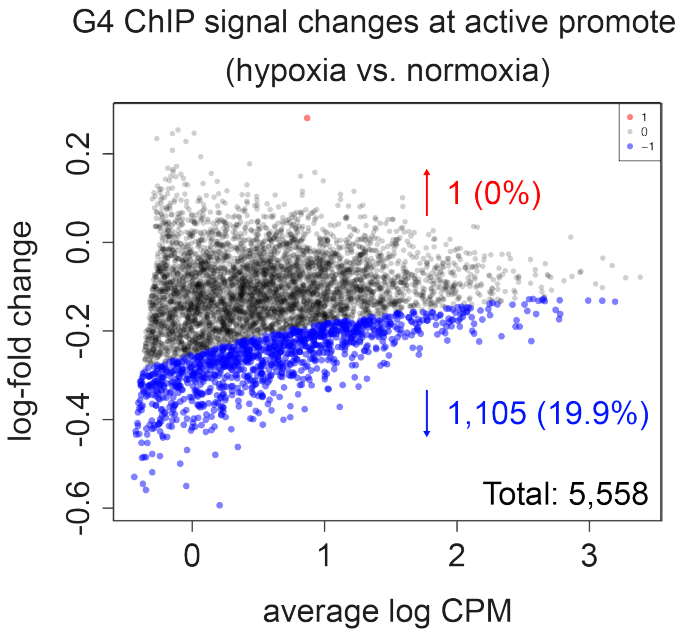

Figure 3.13: Promoter G4s formation is sensitive to chromatin compaction.
MA plot showing fold change in G4 ChIP-seq signal following hypoxia at active gene (Pol II$^+$ G4$^+$) promoters. Blue and red represent sites with significantly ($p < 0.05$) decreased and increased signal respectively. Black represents sites with no changed signal intensity. CPM, read counts per million. Normoxia refers to cells cultured in 21% O_2 and hypoxia refers to cells exposed to 1% O_2 for 1 h. Virtually, no new sites are gained under hypoxia, whereas almost 20% lose signal intensity is found upon hypoxia treatment.

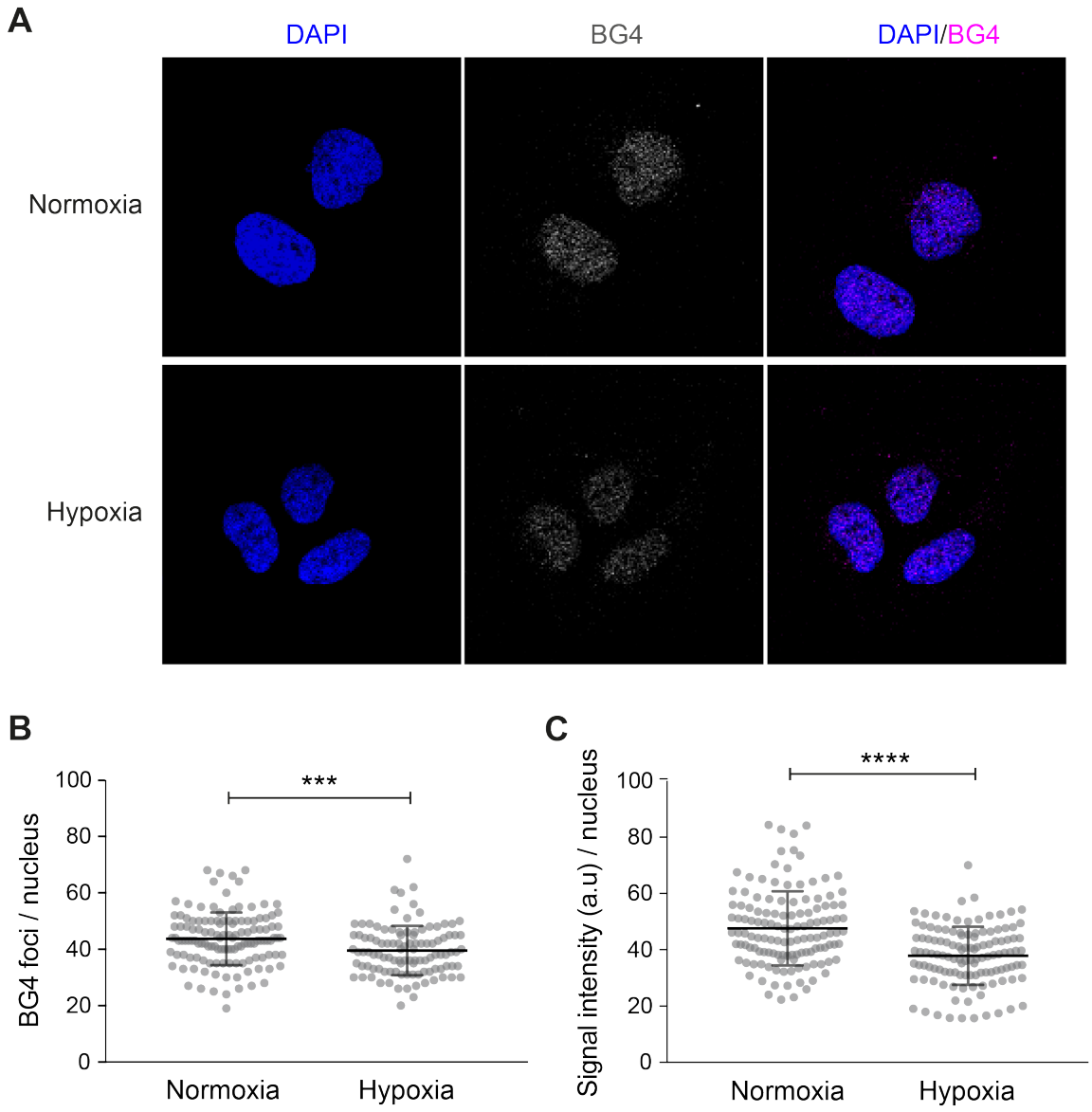

Figure 3.14: BG4 immunofluorescence staining for G4 foci in normoxic versus hypoxic cells.
U2OS cells were fixed in 4% paraformaldehyde, treated with RNase A and stained with the structure-specific BG4 antibody. Confocal images were captured using a Leica SP5 microscope. **(A)** Representative image of BG4 staining (grey/red) in the nuclei (DAPI, blue) of U2OS cells under normoxia or hypoxia. **(B)** Graph quantifying the number of BG4 foci identified in each nucleus. **(C)** Graph showing the average BG4 signal intensity in individual cells (a.u. = arbitrary unit). In panels B and C, each dot represents a single nucleus. 100-150 nuclei were counted per condition and the Mann-Whitney test was performed for each condition across the mean values. Error bars represent the s.d. (n = 1). *** = $p < 0.001$; **** = $p < 0.0001$. In all panels, normoxia refers to cells cultured in 21% O_2 and hypoxia refers to cells exposed to 1% O_2 for 1 h.

As discussed in section 3.3.3, activation of APE1 in response to DNA damage is speculated to result in alterations in cellular G4s, and abrogation of APE1 activity causes the loss of G4s (Roychoudhury et al., 2020). Chronic hypoxia (i.e. cells exposed to hypoxia for more than 24 hours) has previously been demonstrated to cause a reduction in APE1 protein levels (Chan et al., 2014). To evaluate whether the loss of G4s following acute hypoxia could be due to alterations in APE1 levels or activity, protein levels were assessed by Western blotting (Figure 3.15). As before, methoxyamine and H_2O_2 treatments were used as negative and positive controls. Very low levels of total APE1 were seen following H_2O_2 treatment and is consistent with the observation that this induces ubiquitin-mediated degradation of APE1 (Busso et al., 2009). H_2O_2 treatment resulted in increased levels of active APE1 (i.e. acetylated APE1) as signified by increased band intensity and increased molecular weight (upward band shift) due to ubiquitination of APE1. The standard ß-actin loading control also showed an upward band shift following H_2O_2 treatment due to its polyubiquitination (Reeg et al., 2020). As expected, methoxyamine treatment resulted in a reduction in total acAPE1 when compared to normoxic cells. Importantly, no alterations in levels of total APE1 or acAPE1 were observed following hypoxia, indicating that the depletion of G4s during hypoxia cannot be the result of alterations of APE1 levels or activity.

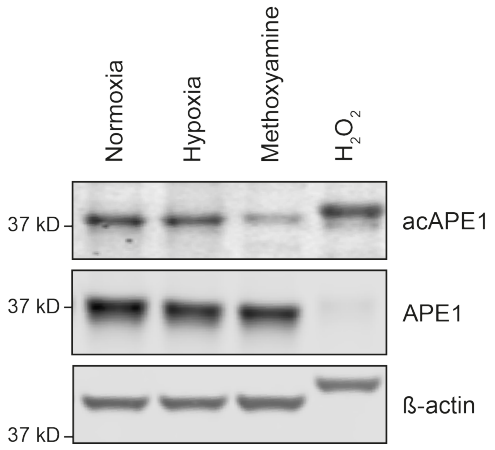

Figure 3.15: Hypoxia does not alter total or active APE1 levels.
Western blotting for APE1 and acAPE1 in K562 cells under normoxic or hypoxic conditions. Methoxyamine and H_2O_2 treated cells represent negative and positive controls respectively. ß-actin is the loading control.

3.3.7 Loss of promoter G4s upon chromatin compaction correlates with loss of RNA Pol II occupancy

Having established that chromatin compaction causes a loss of endogenous promoter G4 signal, I next assessed the impact on RNA Pol II occupancy using Pol II ChIP-seq. As discussed in Chapter 2 (see section 2.3.3; Figure 2.15), G4 containing promoters show greater polymerase occupancy as opposed to their non-G4 marked counterparts. Hypoxia was found to cause a global reduction of RNA Pol II occupancy at promoters (Figure 3.16A). This is in agreement with previous reports of hypoxia-induced transcriptional repression (Batie et al., 2018; Rocha, 2007). Decreased RNA Pol II occupancy was found at the promoters of genes that were previously shown to have reduced gene expression following hypoxia in K562 cells (Lorenzini et al., 2019). This is exemplified by the loss of RNA Pol II occupancy at *VIM* and *IGF2BP2* promoters following hypoxia (Figure 3.16B). Notably, promoters that are marked with a G4 ChIP peak and an RNA Pol II ChIP peak (i.e. Pol II^+ $G4^+$) showed a far greater reduction in RNA Pol II occupancy when compared to Pol II^+ $G4^-$ promoters (Figure 3.16A).

To statistically test the differences in RNA Pol II occupancy, Pol II^+ $G4^+$ promoters with 2-fold or more Pol II ChIP signal compared to corresponding inputs were selected and compared before and after acute hypoxia treatment. Of the 8,307 Pol II^+ $G4^+$ promoters, 1348 showed a significant reduction in Pol II signal intensity ($p < 0.05$; 1,348/8,307 down, 16.2%; Figure 3.17), whereas a negligible number of sites (2/8307) showed increased Pol II signal intensity following hypoxia. G4s therefore mark promoters with elevated RNA Pol II occupancy and hypoxia induced chromatin compaction causes a general loss of G4 signal and RNA Pol II occupancy. This raises the question whether the loss of promoter G4s was associated with the loss of RNA Pol II. To assess the relationship between G4 and Pol II loss following hypoxia, the co-occurrence of changes in G4 signal and RNA Pol II signal was analysed. The overlap between promoters that show increased or decreased ChIP signals for G4 or for Pol II was plotted as percentage of co-occurrence (Figure 3.18A). Interestingly, reduced RNA Pol II occupancy occurred almost entirely at sites that also showed decreased G4 signal. These results demonstrate an association between loss of G4s and reduction in Pol II occupancy upon chromatin compaction. This is exemplified by genome browser views for the *EIF4E-BP1* and *KCTD5* gene promoters (Figure 3.18B).

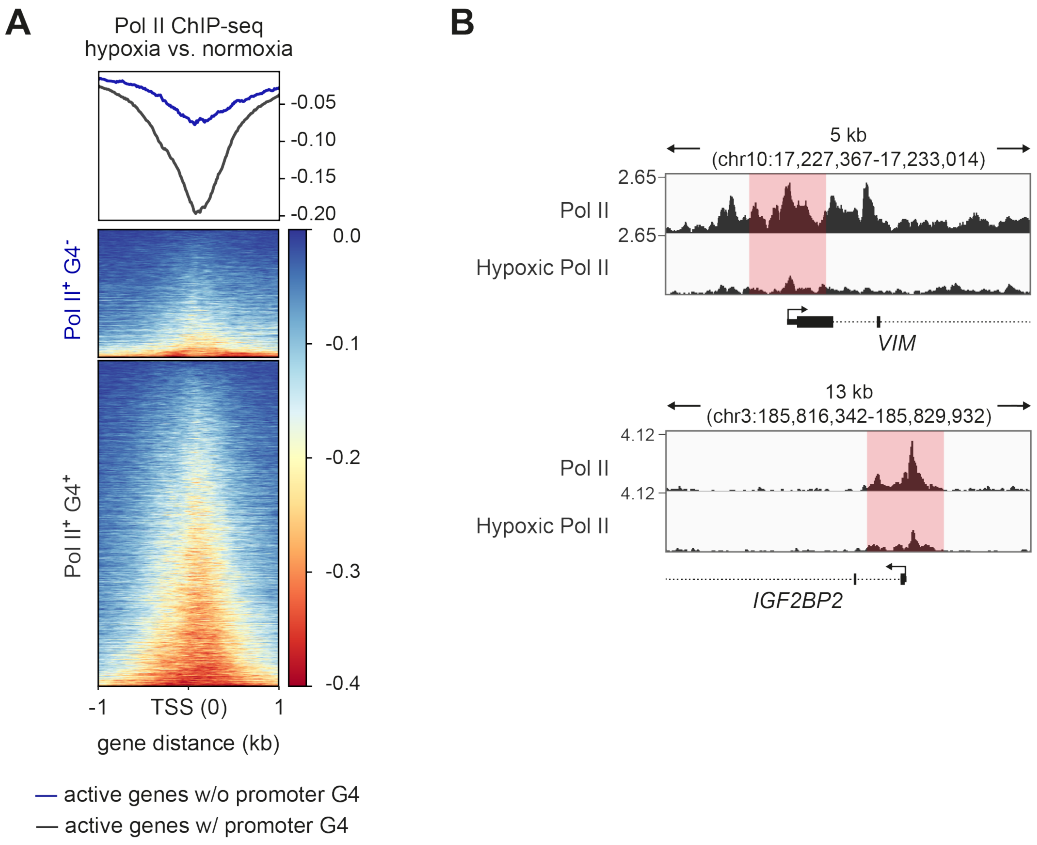

Figure 3.16: RNA Pol II occupancy is reduced following hypoxia treatment.
(A) Signal differences in Pol II ChIP-seq signal for active genes with (black) or without (blue) a promoter G4 in hypoxia versus normoxia. Top panel, metagene plot of the normalized Pol II ChIP-seq signal centered at the TSS. Bottom panel, data plotted for individual loci and represented by a heatmap. RNA Pol II loading at Pol II^{+} G4^{+} promoters shows greater reduction upon hypoxia when compared to Pol II^{+} G4^{-} promoters. **(B)** Example genome browser view showing significantly reduced Pol II occupancy at promoters (red masks) for cells cultured under hypoxic versus normoxic conditions. Genomic coordinates for *VIM* and *IGF2BP2* are indicated. In all panels normoxia refers to cells cultured in 21% O_2 and hypoxia refers to cells exposed to 1% O_2 for 1 h.

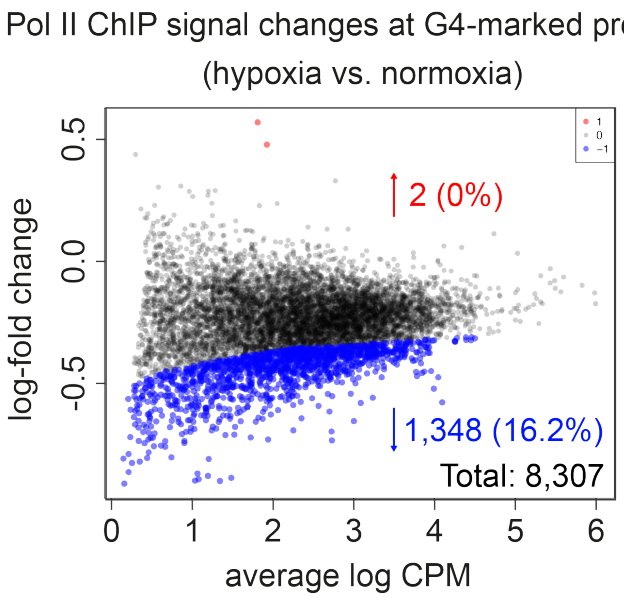

Figure 3.17: The occupancy of RNA Pol II occupancy at G4 marked promoters is reduced following hypoxia-induced chromatin compaction.

MA plot showing fold change in G4 ChIP-seq signal following hypoxia at active gene promoters. Blue and red, sites with significantly ($p < 0.05$) decreased and increased signal respectively. Black represents sites with no changed G4 ChIP-seq signal. CPM, read counts per million. Normoxia refers to cells cultured in 21% O_2 and hypoxia refers to cells exposed to 1% O_2 for 1 h. This indicates that many G4-marked promoters lose RNA Pol II loading upon hypoxia.

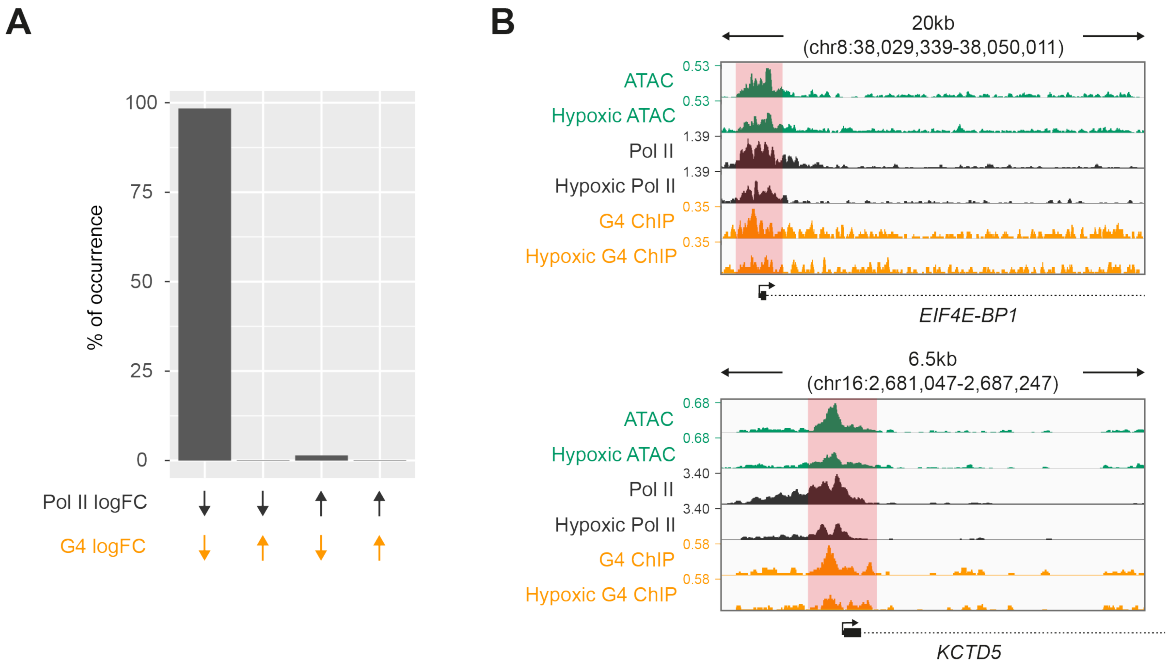

Figure 3.18: Promoter G4 loss correlates with Pol II loss following chromatin compaction.
(A) Co-occurrence of the loss of G4s and RNA Pol II occupancy at promoters under hypoxia. Genomic overlap of promoter G4s that have increases or reductions in G4 ChIP signal (G4 logFC + or - respectively) with increases of reductions in RNA Pol II occupancy (Pol II logFC + or - respectively) at the same site plotted as percentage of co-occurrence. This suggests that reduced RNA Pol II occupancy occurs amongst entirely at sites that also show reduced G4 ChIP-seq signal. **(B)** Genome browser view displaying examples of promoters (e.g. *EIF4E-BP1* and *KCTD5*) that show significant reductions in ATAC-seq (green), Pol II ChIP-seq (black) and G4 ChIP-seq peaks (yellow) for cells under normoxic (top) and hypoxic (bottom) conditions. Genomic coordinates are indicated above. This demonstrates that loss of G4s upon chromatin compaction is accompanied with reduction in Pol II occupancy.

3.4 Summary

In this chapter, I investigated the interaction between chromatin structure and promoter G4s in cells by assessing the impact of reduced chromatin accessibility on G4 formation. To induce chromatin compaction in cells, but minimise possible downstream off-target effects on promoter G4 folding, acute chemical treatments targeting histone acetyltransferases were used. Despite treatment with a very high dose (100uM), the majority of the HAT inhibitors tested did not significantly diminish levels of histone acetylation within 4 hours. Multiple studies have reported that these compounds, at doses similar to what was tested in this chapter, cause inhibition of cell growth following days of treatment (Park et al., 2018; Trisciuoglio et al., 2012; Zhao et al., 2019). Chemical properties of HAT inhibitors, such as poor permeability (Dahlin et al., 2017), may be one reason why such long treatment may be required. However, since these compounds are also often associated with non-specific downstream effects on cellular response, increased treatment time was considered undesirable.

Amongst the inhibitors tested, a high concentration of garcinol caused rapid cell death within 4 h, which is likely to be due to the induction of apoptosis. In human leukemia cells, garcinol causes release of cytochrome c from mitochondria, activation of caspase 3 and subsequent cell apoptosis (Pan et al., 2001). Garcinol can also result in apoptosis by impeding the survival signalling pathway in cells (Liao et al., 2005) or by stimulating ROS formation and energy depletion (Fazio et al., 2015). To minimise compounding effects, I treated cells with a lower dose of garcinol (5 µM for 4 h) and was able to detected substantial reduction of histone acetylation levels. However, even at this dose, garcinol induced cellular DNA damage response pathways, as evidenced by elevated acetylation of APE1, which marks increased APE1 activity and induction of the base excision repair pathway. Multiple studies have linked G4s with DNA damage (Fleming et al., 2017; Fouquerel et al., 2019), with increased APE1 activity suggested to increase G4 stabilisation (Roychoudhury et al., 2020). It is therefore difficult to draw mechanistic links between chromatin compaction and promoter G4s by using garcinol or any chemical agent that elicits DNA damage.

Acute hypoxia was used as an alternative method to induce chromatin compaction. Importantly, to minimise the possibility that cellular response other than chromatin compaction affect G4 landscape, cells were exposed to hypoxia for only 1 hour. In comparison to garcinol, hypoxia led to a more robust genome-wide reduction in chromatin accessibility, as seen by MNase digestion and ATAC-seq fragment size measurements. Moreover, hypoxia caused significant compaction at greater than 85% promoter G4s as opposed to ~24% seen after garcinol treatment. In addition, acute hypoxia (1% O_2, 1 h) is not known to be associated with DNA damage, as evidenced by the lack of APE1 activation. Acute hypoxia thus provides a better model system for rapid induction of chromatin compaction in K562 cells.

A fifth of promoter G4s showed significantly decreased G4 ChIP signal upon hypoxia, indicating that many promoter G4s are sensitive to chromatin compaction. Furthermore, the loss of promoter G4 signal was accompanied with loss of RNA Pol II occupancy at these sites, demonstrating an association between loss of G4s and reduction in Pol II occupancy. Data in Chapter 2 clearly demonstrate that the reduction in transcriptional activity does not lead to a loss of G4s, thus the loss of promoter G4s in this chapter can be attributed to chromatin compaction. However, the detailed mechanism of how G4s become unfolded during chromatin compaction following hypoxia remains unknown.

To sum up, whilst Chapter 2 demonstrated that the folding of promoter G4s in cells precedes active transcription, the work in Chapter 3 suggests that G4 formation in promoters is facilitated by an accessible chromatin environment. Moreover, these data highlight a strong correlation between the loss of promoter G4s and the loss of Pol II occupancy in hypoxia, which may indicate a direct link between certain G4s and polymerase recruitment. This concept will be further investigated in Chapter 4.

Chapter 4

Identifying the role of promoter G4s in transcription

4.1 Background

The work in Chapter 3 described studies that showed chromatin compaction leads to the loss of signal at many promoter G4s, suggesting a causative link between formation of certain promoter G4s and accessible chromatin. Moreover, it was also found that promoter G4s loss is concomitant with loss of RNA Pol II occupancy upon chromatin compaction. Given that the formation of promoter G4s precedes active transcription (Chapter 2), it is plausible that upon chromatin compaction, RNA Pol II is lost as a result of the loss of G4s.

As previously reported (see section 2.1), binding motifs for several zinc-finger transcription factors are strongly enriched in potential G4-forming motifs (Kumar et al., 2011; Todd and Neidle, 2008). Moreover, endogenous G4 structures observed with G4 ChIP-seq overlap with binding sites for many transcription factors (Hou et al., 2019), suggesting that promoter G4s could recruit TFs. In support of these findings, several independent studies have reported *in vitro* interactions of G4-structured oligonucleotides with TFs, such as SP1, XPB, CNBP and LARK (Borgognone et al., 2010; Gray et al., 2014; Niu et al., 2019; Raiber et al., 2012). The binding affinity of TFs to G4s has in a number of cases been found to be comparable to their affinities to consensus double-stranded binding sequences (Cogoi et al., 2010; Raiber et al., 2012). Furthermore, G4s may directly recruit RNA Pol II via interactions with polymerase subunits. Mass spectrometric proteomics analysis of proteins specifically pulled down by a G4 forming oligonucleotide found subunits of RNA Pol II to specifically associate with folded G4s (Makowski et al., 2018). Together, these studies indicate a role for promoter G4s in recruitment of RNA Pol II.

Since sites that lose promoter G4 signal upon chromatin compaction were found to have preferentially reduced Pol II occupancy, I hypothesised that artificial stabilisation of G4 structures at these sites may lead to retention of RNA Pol II binding. Furthermore, since G4s are found predominantly in accessible chromatin regions (Hänsel-Hertsch et al., 2016; Kouzine et al., 2017; Wong and Huppert, 2009), it is possible that stabilising G4s may lead to

changes in chromatin compaction following hypoxia, as ligand bound G4s may present a physical impediment to chromatinisation processes.

Small molecules that specifically bind G4 structures have been used for artificially stabilising G4s in cells. Such ligands have often been used for studying the function of G4 structures in modulating gene expression (see section 1.6.1). For instance, stabilisation of G4s using the G4 ligand TMPyP4 showed inhibitory effects on transcriptional activity of *MYC* and *KRAS* (Cogoi and Xodo, 2006; Siddiqui-Jain et al., 2002). Isoalloxazine ligands, which also stabilise G4s, were found to cause a substantial reduction of *KIT* expression in cells (Bejugam et al., 2007). Furthermore, the G4 binding ligand PDS was shown to target genes containing G4 motifs and PDS treatment resulted in reduced expression of these targeted genes (Rodriguez et al., 2012). Subsequently, cells treated with PDS also showed a global increase of G4 formation as seen by increased formation of G4 foci with BG4 immunofluorescence (Biffi et al., 2013b). However, various studies have shown that treatment of cells with many G4 stabilising ligands for more than 12 hours can lead to DNA damage (Fleming et al., 2017; Fouquerel et al., 2019; Rodriguez et al., 2012). The mechanism of how G4 ligands cause DNA damage is not entirely clear, but stabilised G4s are thought to interfere with cellular processes such as transcription and DNA replication, thus resulting in strand breakage and initiation of a DNA damage response.

Alternative approaches to modulate G4 formation have also been used to investigate interplay between G4s and transcriptional activity (see section 1.6.2). For example, the introduction of constructed plasmids, which contains a G4 motif on the template strand downstream of the promoter of a luciferase reporter gene, in human embryonic kidney cells resulted in a substantial inhibitory effect on luciferase expression (Agarwal et al., 2014). By contrast, transcription was not impacted when the G4 motif was on the non-template strand. Although such models may provide information regarding site-specific and strand-specific influence of G4s on transcription, they are limited by the lack of appropriate chromatin context. Such limitations may be significant since the findings of my studies, described thus far in this thesis, have shown that endogenous G4 formation is highly dependent on the chromatin context.

In this chapter, I test the hypothesis that prior stabilisation of G4s may diminish the influence of acute hypoxia by limiting chromatin compaction and RNA Pol II loss. To achieve genome-wide G4-stabilisation in cells, I treated K562 cells with the small molecule pyrrolidine PDS (pyPDS) (Müller et al., 2010). pyPDS has recently been conjugated with a fluorescent probe to adapt it to the visualisation of native G4 structures in living cells (Di Antonio et al., 2020), corroborating the binding of pyPDS to G4 structures in cells. Furthermore, to minimise downstream cytotoxic effects, particularly DNA damage responses, associated with such G4-stabilising ligands, short treatment times were used.

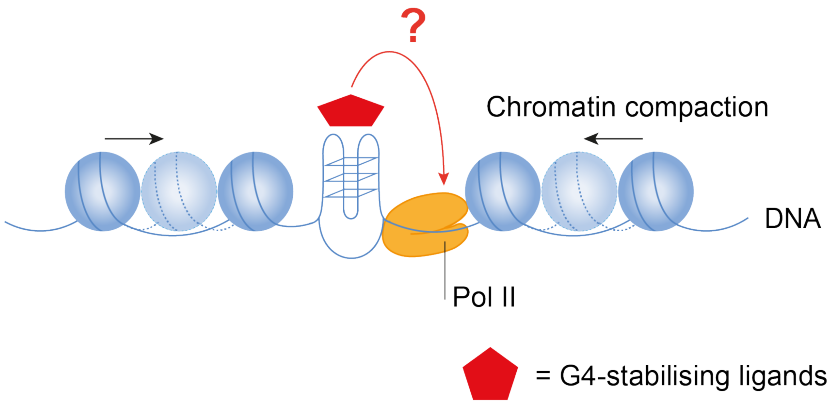

Figure 4.1: Investigating the role of stabilised promoter G4s on chromatin compaction and RNA Pol II occupancy following acute hypoxia.
Graphical representation of experimental design studying how stabilisation of G4s by small molecules alters chromatin structure and RNA Pol II loading under hypoxia-driven chromatin compaction. Cells were treated with G4-stabilising ligands prior to the exposure to hypoxia. The impact of pre-stabilised G4s on chromatin accessibility and RNA Pol II occupancy upon hypoxia was assessed by ATAC-seq and RNA Pol II ChIP-seq.

4.2 Objectives

The aim of this chapter is to investigate how G4 ligand-mediated stabilisation impacts hypoxia driven chromatin compaction and loss of RNA Pol II at promoter G4s. The first objective is to examine the G4-stabilising capability of pyPDS in cells using BG4 immunofluorescence and G4 ChIP-seq. The second objective is to assess if pyPDS could stabilise G4s in cells in hypoxia-induced chromatin compaction. The third objective is to assess the impact of G4 stabilisation on chromatin accessibility upon hypoxia. The final objective of this chapter was to investigate the influence of stabilised G4s on RNA Pol II occupancy following exposure to acute hypoxia.

4.3 Results

4.3.1 G4 ChIP-Seq is unsuitable to detecting ligand-stabilised G4s in cells

To directly assess the ability of pyPDS to increase the stability of promoter G4s in cells, G4 ChIP-seq was performed on chromatin from cells treated with 1 μM pyPDS or DMSO control for 1 hour. G4 ChIP-seq on three independent biological replicates, each with three technical replicates were performed. Consensus promoter G4 peaks were defined and filtered as described previously (see section 3.3.5 and section 3.3.6). Active gene promoters were defined using RNA Pol II occupancy from RNA Pol II ChIP-seq. Pol II^+ $G4^+$ sites were then specifically selected to compare G4 formation in the absence or presence of pyPDS.

It would be expected that stabilisation of G4s by pyPDS would result in increases in G4 ChIP signals. However following pyPDS treatment, it was striking that of 5,912 Pol II^+ $G4^+$ promoter sites assessed, nearly 70% showed a statistically significant reduction in signal intensity ($p < 0.05$; 4,046/5,912, 68.4%; Figure 4.2A), whereas a negligible number of sites (1/5,912) showed increased G4 ChIP signal intensity. The reduction of G4 ChIP signal was also found at genes where promoter G4s have been previously suggested to be stabilised by G4-stabilising ligands, for example *MYC* promoter and *VEGF* promoters (Siddiqui-Jain et al., 2002; Sun et al., 2011) (Figure 4.2B). These findings indicate that treatment with the G4 stabilising ligand pyPDS results in the loss of G4 ChIP signal at promoters and is contrary to my expectation.

There may be a different explanation for the observed signal loss in these experiments. Many studies have shown that stable G4s can act as obstacles to DNA synthesis by stalling DNA polymerases (Han et al., 1999; Woodford et al., 1994). Indeed, this has been utilised to define genome-wide G4 formation using polymerase stalling by ligand-mediated stabilisation of G4s (Chambers et al., 2015; Dexheimer et al., 2006; Qin et al., 2007). Possibly, the decrease of G4 ChIP signal following pyPDS stabilisation is related to this reduced amplification of G4s in the presence of G4-stabilising ligand during library preparation. As an alternative approach, the next section makes use of BG4 IF staining as this does not require any DNA amplification steps.

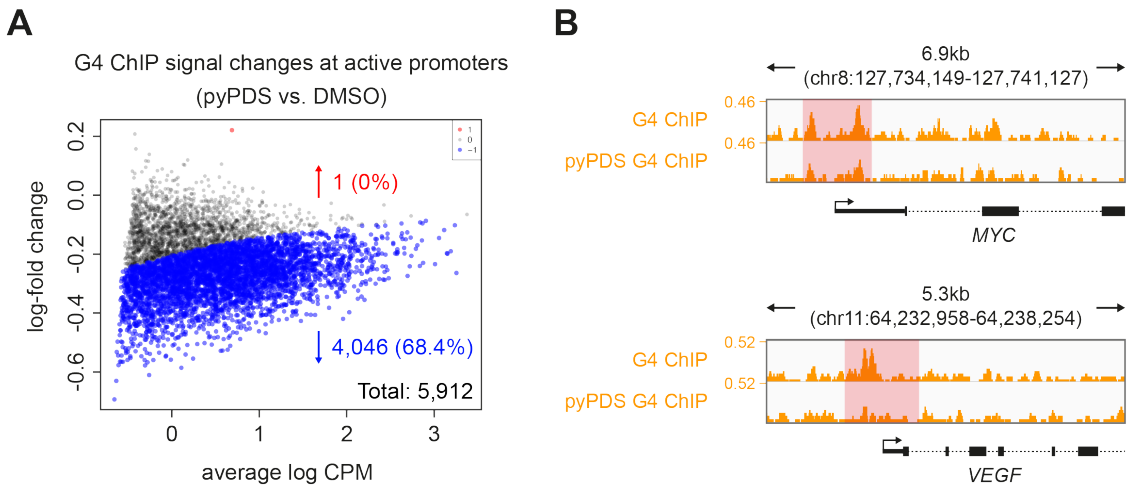

Figure 4.2: Reduction of G4 ChIP-seq signal following pyPDS treatment.
(A) MA plot showing the fold change in G4 ChIP-seq signal at active promoters (Pol II^{+} G4^{+} promoters; $p < 0.05$) in pyPDS-treated versus DMSO-treated K562 cells. Statistically significantly ($p < 0.05$) higher and lower signals are in red and blue respectively; black dots indicate unchanged regions. CPM, read counts per million. This shows that ~20% of sites assessed lose G4 ChIP signal. **(B)** Genome browser view of *MYC* and *VEGFB* exemplify genes that significantly change in G4 ChIP-seq peaks (yellow). Tracks compare peaks between DMSO (0.1%, 2 h) and pyPDS (1 µM, 2 h) conditions. Genomic coordinates indicate track range. This shows reduced G4 ChIP signal at the promoters (masked in red) after pyPDS treatment.

4.3.2 G4 stabilisation protects G4s from unfolding in hypoxia

To assess G4 formation by an alternative approach, nuclear G4s were visualised using BG4 IF staining. Previously this method was originally used to detect an increase in G4 formation upon PDS treatment (Biffi et al., 2013b) and has been widely used to study G4 formation in cells (Biffi et al., 2013a; Hänsel-Hertsch et al., 2016; Liu et al., 2016; Zhang et al., 2018). BG4 IF was performed in U2OS cells since K562 cells are not amenable to G4 imaging. Changes in the number and intensity of BG4 foci were measured (Figure 4.3). Unlike G4 ChIP-seq, BG4 IF is unable to provide positional information regarding specific G4 sites. An average of 48 BG4 foci were identified following pyPDS treatment (1 µM, 2 h), which represents a significant increase ($p < 0.001$; over 150 nuclei from 1 biological replicate) when compared to the average of 43 BG4 foci in DMSO-treated control cells (Figure 4.3B). However, this is not a significant increase in signal intensity (Figure 4.3C). Previously, a 2.9-fold increase of BG4 foci was observed following 24 h treatment of U2OS cells with 10 µM PDS (Biffi et al., 2013b). Comparatively, the increase in BG4 foci with pyPDS I detected is modest, and may

be because cells were treated with a lower dose of pyPDS (1 µM) for a much shorter time (2 h).

The results, which quantified from more than 150 individual nuclei, showed that pyPDS treatment at low doses and times leads to modest but statistically significant increase in cellular G4s. This was sufficiently robust to suggest that pyPDS leads to stabilisation of cellular G4s. I therefore proceed to the next step to investigate whether prior stabilisation by pyPDS protects G4 folding upon subsequent exposure to acute-hypoxia. For this, cells were pre-treated with 1 µM pyPDS for an hour and subsequently exposed to hypoxia or maintained them in normoxia for an additional hour in presence of pyPDS. Cells pre-treated with DMSO that were maintained in normoxia or exposed to hypoxia served as controls. The impact on G4 folding was then measured using BG4 immunofluorescence. Comparable to my observations in Chapter 3 (see section 3.3.6; Figure 3.14), only 30 BG4 foci were identified in DMSO-treated cells in hypoxia, suggesting a significant decrease of G4s upon hypoxia treatment ($p < 0.0001$; Normoxia-DMSO: 43 vs Hypoxia-DMSO: 30; Figure 4.3B). In addition, a significant reduction of BG4 signal intensity was also observed in DMSO-treated cells in hypoxia when compared to normoxia ($p < 0.0001$; Normoxia-DMSO: 24 vs Hypoxia-DMSO: 21; Figure 4.3C). These results confirm that DMSO pre-treatment does not alter the cellular G4 response to hypoxia.

An average of 38 BG4 foci were identified in pyPDS pre-treated hypoxic cells, which shows a significant ($p < 0.001$) increase when compared to the average of 30 BG4 foci in DMSO-treated hypoxic cells (Figure 4.3B). BG4 signal intensity also significantly increased in pyPDS pre-treated hypoxic cells versus DMSO-treated hypoxic cells ($p < 0.001$; Hypoxia-pyPDS: 25 vs Hypoxia-DMSO: 21; Figure 4.3C) These results suggest that pre-treatment with pyPDS is able to diminish the influence of hypoxia on G4s.

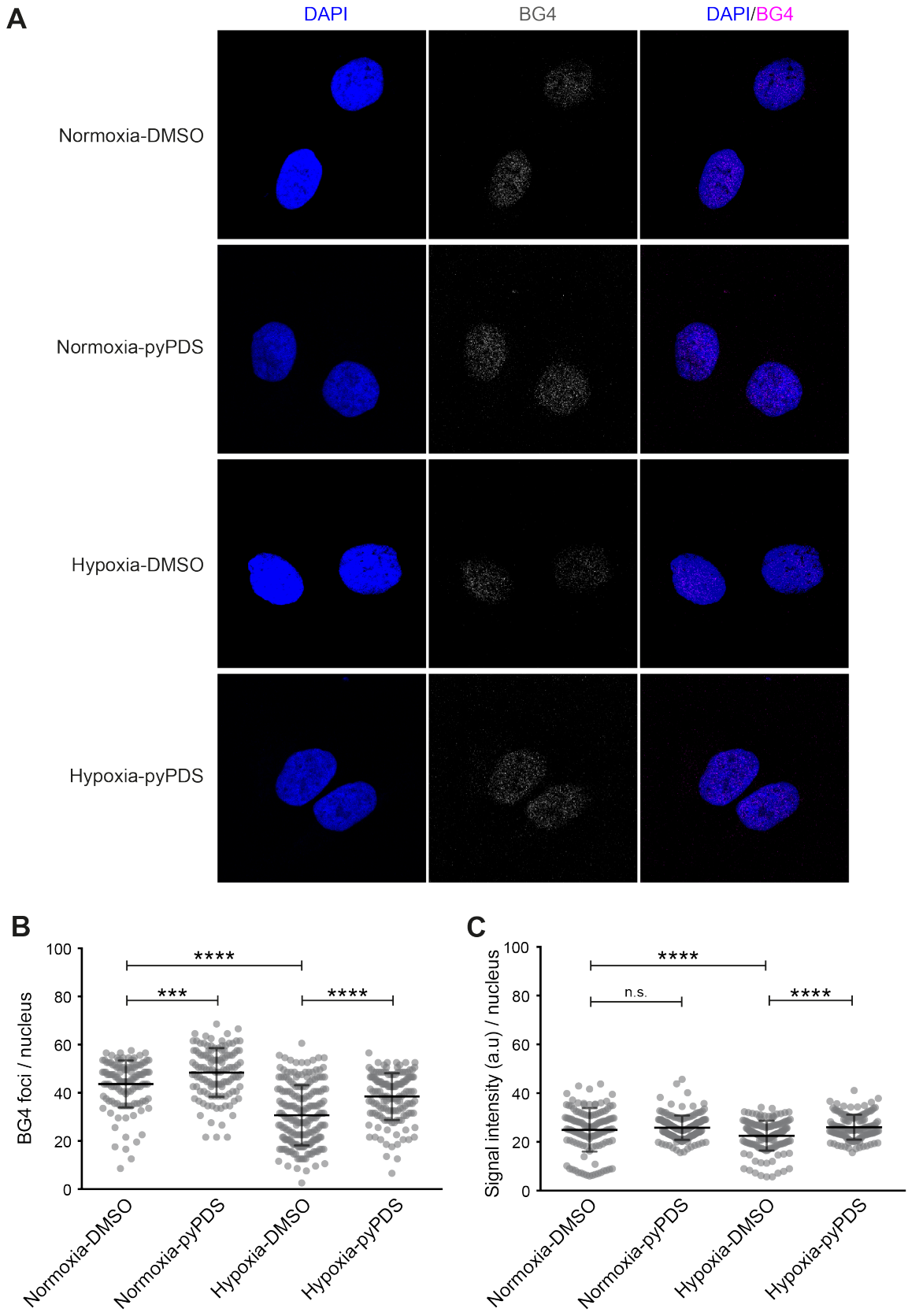

Figure 4.3: pyPDS treatment diminishes the impact of hypoxia on cellular G4s.
U2OS cells were paraformaldehyde fixed, RNase A treated and stained with BG4 antibody. Confocal images were captured using a Leica SP5 microscope. **(A)** BG4 staining (grey/red) in the nuclei stained by DAPI (blue) of DMSO- or pyPDS-treated U2OS cells under normoxia or hypoxia. **(B)** Graph quantifying the number of BG4 foci

identified in each nucleus. Normoxia-DMSO = 43; Normoxia-pyPDS = 48; Hypoxia-DMSO = 30; Hypoxia-pyPDS = 38. **(C)** Graph showing the average BG4 signal intensity in individual cells (a.u. = arbitrary unit). The mean value for each condition is: Normoxia-DMSO = 24; Normoxia-pyPDS = 25; Hypoxia-DMSO = 21; Hypoxia-pyPDS = 25. In panel B and C, each dot represents a single nucleus. An average of 150 nuclei were counted per condition and the one-way ANOVA test was performed for each condition across the mean values. Error bars represent the s.d. *** = $p < 0.001$; **** = $p < 0.0001$. In all panels, normoxia refers to cells cultured in 21% O_2 and hypoxia refers to cells exposed to 1% O_2 for 1 h. These results suggest that pyPDS-treated cells show more BG4 foci than DMSO-treated cells in normoxia. Less BG4 foci and reduced G4 signal intensity are found in DMSO-treated cells following hypoxia when compared to normoxia. Furthermore, the increase of BG4 foci and G4 signal intensity are found in pyPDS pre-treated hypoxic cells relative to DMSO pre-treated hypoxic cells.

As discussed previously (see section 3.3.3), G4 formation may be enhanced by cellular responses to DNA damage, especially by BER-associated APE1 activation (Roychoudhury et al., 2020). As mentioned previously, treatment of cells with G4-stabilising ligands is often associated with activation of DNA damage response (Fleming et al., 2017; Fouquerel et al., 2019; Rodriguez et al., 2012), which I attempted to minimise by choosing very short treatment times. However, it is entirely reasonable to challenge that the increase of BG4 foci following pyPDS treatment occurs indirectly via induction of BER and activation of APE1. To examine this, Western blotting was performed to assess APE1 activation upon pyPDS treatment in hypoxic or normoxic conditions (Figure 4.4). As for prior experiments (see section 3.3.3 and section 3.3.6), methoxyamine and H_2O_2 treatments were used as negative and positive controls. Very low levels of total APE1 could be observed following H_2O_2 treatment due to the induction of ubiquitin-mediated degradation of APE1 (Busso et al., 2009). Methoxyamine treatment resulted in a reduction in acAPE1, while H_2O_2 caused an upward band shift due to ubiquitination of APE1. The standard ß-actin loading control also showed an upward band shift following H_2O_2 treatment due to its polyubiquitination (Reeg et al., 2020). However, normoxic and hypoxic cells pre-treated with pyPDS did not demonstrate any increase in levels of total or active APE1 (i.e. acetylated APE1) when compared to DMSO pre-treated control cells. Therefore, the induction of G4s upon pyPDS treatment is not due to alterations in APE1 levels or activity.

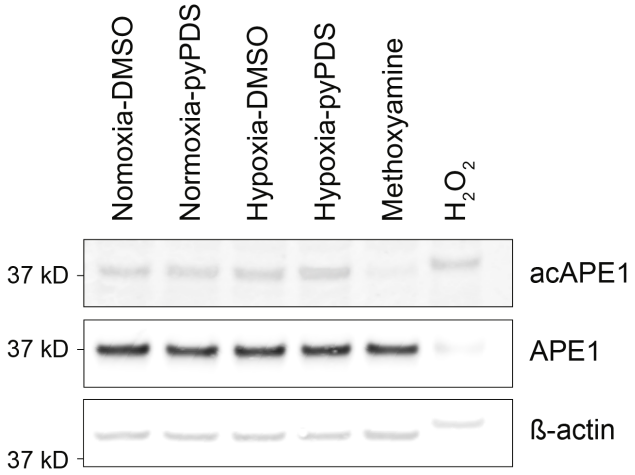

Figure 4.4: Addition of pyPDS does not increase APE1 level and its activity.
Western blotting for APE1 and acAPE1 in DMSO- or pyPDS-treated cells upon hypoxia or normoxia. The acAPE1 protein level suggests the activity of APE1 binding and repair. Methoxyamine and H_2O_2 treated cells represent negative and positive controls respectively. ß-actin serves the loading control. This shows that pyPDS and hypoxia treatment does not affect levels of total or active APE1.

4.3.3 pyPDS treatment does not alter chromatin accessibility and RNA Pol II recruitment at promoter G4s

Prior to investigating the impact of G4 stabilisation on chromatin accessibility and Pol II occupancy at promoters in response to hypoxia, it was important to confirm that pyPDS had no effects on chromatin accessibility and RNA Pol II occupancy in normoxia. For this, ATAC-seq and RNA Pol II ChIP-seq were performed in normoxic K562 cells treated with DMSO or pyPDS. To examine the impact of G4 stabilisation on chromatin, changes in ATAC-seq signal at G4-marked active gene promoters (i.e. Pol II^+ $G4^+$) were assessed. Of 9,217 sites analysed, negligible alterations in ATAC-seq signal were observed in pyPDS treated cells when compared to DMSO ($p < 0.05$; 3/9,217 up and 2/9,217 down; Figure 4.5). This indicates that short-term ligand-mediated stabilisation of G4s by pyPDS does not measurably alter general chromatin structure under homeostatic growth conditions.

To assess alterations in Pol II occupancy upon pyPDS treatment, RNA Pol II ChIP signals at Pol II^+ $G4^+$ promoters, which were two-fold enriched over input, were compared between pyPDS- and DMSO-treated cells. As observed for ATAC-seq, no alterations in Pol II

occupancy were observed under normoxic conditions with pyPDS when compared to DMSO ($p < 0.05$; Figure 4.6). This result shows that pyPDS treatment does not significantly alter RNA Pol II occupancy at G4-marked promoters in homeostatic cells. This also rules out the possibility of any confounding effects of pyPDS on chromatin compaction and RNA Pol II occupancy. The data therefore serve as important controls for the experiments to follow.

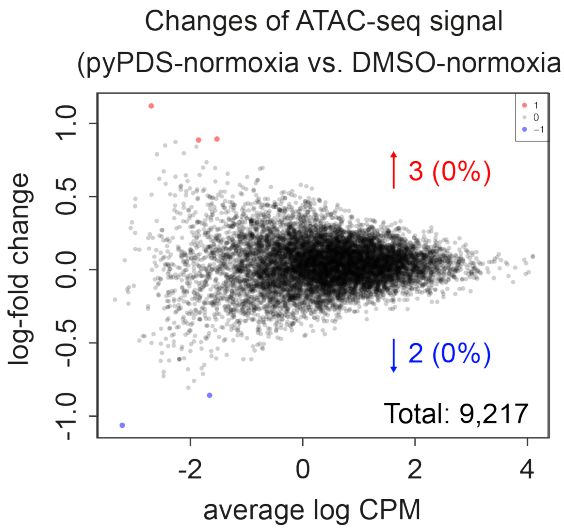

Figure 4.5: pyPDS treatment does not affect chromatin accessibility at G4-marked active promoters under normoxia.

MA plot showing fold change in ATAC-seq signal at Pol II^{+} G4^{+} promoters between pyPDS- and DMSO-treated cells upon normoxia. Significant increases or decreases ($p < 0.05$) in ATAC-seq signal in pyPDS-treated cells relative to DMSO-treated cells are indicated in red and blue respectively; black dots indicate regions that show no change in ATAC-seq signal. CPM, read counts per million.

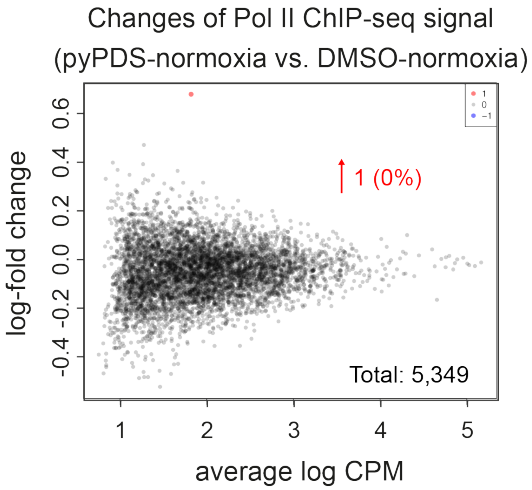

Figure 4.6: pyPDS treatment does not affect RNA Pol II occupancy at G4-marked active promoters under normoxia.
MA plot showing fold change in RNA Pol II ChIP-seq signal at Pol II$^+$ G4$^+$ promoters between pyPDS- and DMSO-treated cells upon normoxia. Significantly increases ($p < 0.05$) in RNA Pol II ChIP-seq signal in pyPDS-treated cells relative to DMSO-treated cells are indicated in red, and black dots indicate unchanged regions. CPM, read counts per million.

4.3.4 Ligand mediated stabilisation of G4s fails to alter hypoxia driven chromatin compaction

Since G4s form predominantly in open chromatin regions (see section 2.3.2; Figure 2.12), I next hypothesised that stabilised G4s may physically impede chromatin compaction in hypoxia. Having provided data to suggest that pre-stabilisation of G4s by pyPDS protects G4s from unfolding upon hypoxia (see section 4.3.2), I evaluated whether this stabilisation of G4s impacts upon chromatin compaction in hypoxic cells. To address this question, the change of chromatin accessibility at promoter regions was assessed by comparing ATAC-seq signal from pyPDS-treated hypoxic K562 cells with ATAC-seq signal from DMSO-treated cells. At G4-marked active promoters (i.e. Pol II$^+$ G4$^+$ promoters), a reduction of ATAC-seq signal was observed in pyPDS-treated hypoxic cells (Figure 4.7), indicating that hypoxia-driven chromatin compaction still ensues at promoter G4s despite pre-stabilisation by pyPDS. Remarkably, this also suggests that stabilisation by pyPDS enables G4s to persist despite chromatin compaction, however this cannot be directly demonstrated due to the previously discussed incompatibility of G4 ChIP-seq with pyPDS.

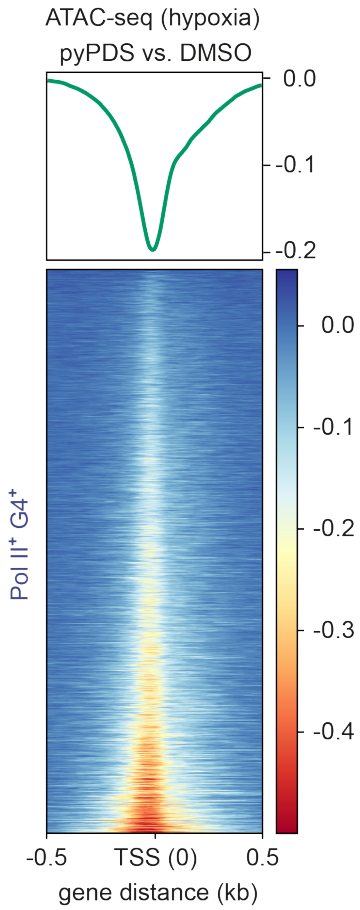

Figure 4.7: Chromatin compaction still ensues in pyPDS-treated hypoxic cells.
Chromatin accessibility differences for G4-marked active promoters (i.e. Pol II$^+$ G4$^+$ promoters) under hypoxia for cells pre-treated with pyPDS or DMSO. Top panel, metagene plot of normalised ATAC signal centered at the TSS showing the signal difference between DMSO- and pyPDS-treated hypoxic cells. Bottom panel, normalised ATAC-seq signal plotted for all individual loci and collectively represented by a heatmap. Hypoxia refers to cells exposed to 1% O_2 for 1 h.

4.3.5 Stabilised G4s override chromatin compaction to recruit RNA Pol II to promoters

Since the loss of promoter G4s upon chromatin compaction coincides with the loss of RNA Pol II at the same sites (see section 3.3.7; Figure 3.18), it is possible that the unfolding of promoter G4s influences RNA Pol II occupancy. The data so far suggests that pyPDS pre-treatment can stabilise cellular G4s, which chromatin compaction still maintains. This allowed me to ask whether the stabilisation of G4s by pyPDS could cause retention of RNA Pol II binding at promoters in spite of chromatin compaction. To address this, polymerase loading at the promoters was assessed by RNA Pol II ChIP-seq, and 5,349 promoter G4 sites that showed significant RNA Pol II ChIP signal enrichment (i.e. sites that show greater than 2-fold Pol II ChIP signal compared to the corresponding input) were analysed. In DMSO-treated cells, about 10% of the Pol II^+ $G4^+$ promoters showed a significant decrease in Pol II ChIP signal following the exposure to hypoxia ($p < 0.05$; 566/5,349, 10.6%; Figure 4.8A), which is comparable to observations in the previous chapter (see section 3.3.7; Figure 3.17). In contrast, only 1.2% of Pol II^+ $G4^+$ sites showed significant reduction of Pol II ChIP signal upon hypoxia when G4s were pre-stabilised by pyPDS ($p < 0.05$; 64/5,349; Figure 4.8B). This shows that the addition of pyPDS reduced the number of sites showing Pol II loss in hypoxia by 10-fold (10.6% vs 1.2%). This is apparent when visualising the overlap in sites that show significant loss of RNA Pol II in DMSO-treated cells upon hypoxia with those that show significant Pol II loss in pyPDS-treated hypoxic cells versus DMSO-treated normoxic cells (Figure 4.8C). When compared to DMSO-treated normoxic cells, hypoxia treatment resulted in the loss of RNA Pol II in 566 promoters, whereas pre-treatment of pyPDS in hypoxia showed Pol II loss in only 64 promoters. With the pre-treatment of pyPDS, RNA Pol II loss in hypoxia was no longer observed at the majority of the promoters (505/566, 89.2%) that were previously shown to have reduced Pol II ChIP signal upon hypoxia. Together, these results demonstrate that pyPDS treatment leads to polymerase retention at a majority of sites that would otherwise be lost in hypoxia.

Furthermore, a substantial increase of RNA Pol II ChIP signal at G4-marked promoters was observed following pyPDS stabilisation when compared to non-stabilised hypoxic conditions (Figure 4.9). This corroborates my findings that pyPDS treatment retains RNA Pol II at

individual promoter G4s in hypoxia. Importantly, no changes in RNA Pol II occupancy were seen at non-G4-marked promoters. This rules out non-specific, G4-independent, ligand-associated effects on RNA Pol II recruitment in hypoxia. Overall, these results suggest that the chemical stabilisation of G4 structures, which would otherwise be lost during chromatin compaction, leads to retention of Pol II in spite of chromatin compaction signals. The findings are exemplified by genome browser views for the *ATXN1L* and *FSCN1* gene promoters (Fig. 4.10).

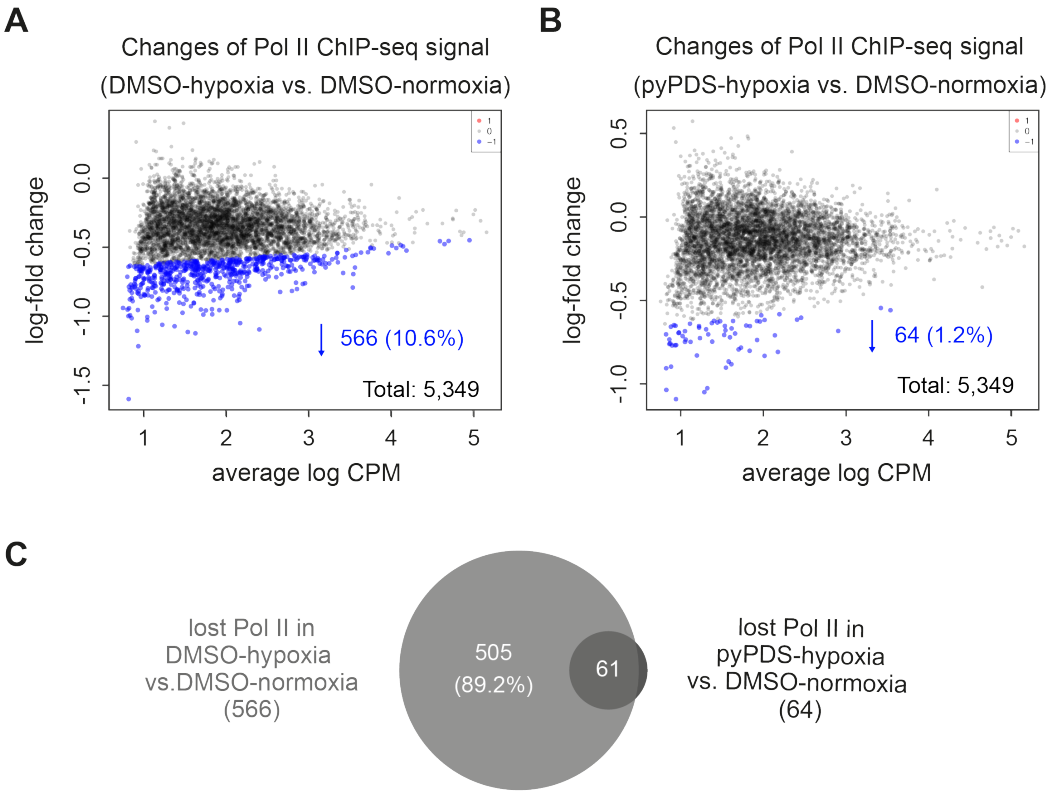

Figure 4.8: pyPDS stabilisation substantially reduces the loss of RNA Pol II sites at G4-marked promoters upon hypoxia.
(A) MA plot showing fold change in RNA Pol II ChIP-seq signal between hypoxic and normoxic cells treated with control DMSO at G4-marked active (i.e. Pol II^{+} G4^{+}) gene promoters. **(B)** As in panel A but for changes in Pol II ChIP-seq signal for cells treated with pyPDS and subjected to hypoxia compared to DMSO-treated cells under normoxia. In both panel A and B, significantly decreased signal ($p < 0.05$) in Pol II ChIP-seq is indicated in blue. black dots indicate sites that have unchanged Pol II ChIP signal. CPM, read counts per million. **(C)** Venn diagram showing Pol II^{+} G4^{+} promoter regions that have lost Pol II occupancy between DMSO-treated normoxic and hypoxic cells and overlap with sites lost in pyPDS-treated hypoxic cells compared to DMSO-treated normoxic conditions. These results demonstrate that the majority of RNA Pol II sites seen reduced in hypoxia are retained by pyPDS.

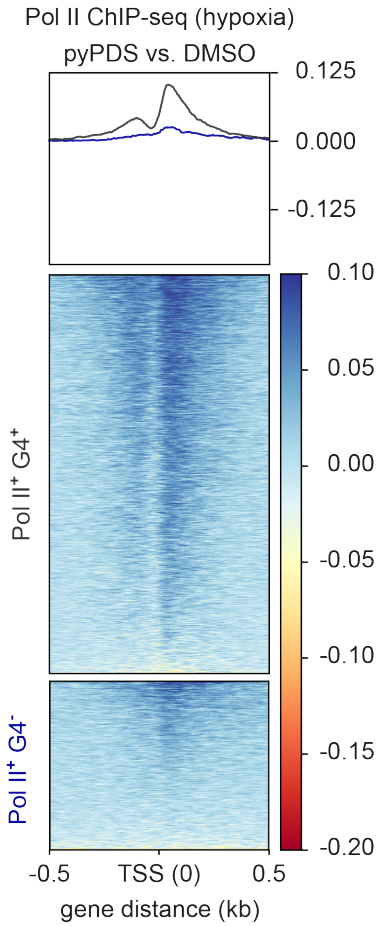

Figure 4.9: Stabilisation of G4s by pyPDS retains RNA Pol II at promoters upon hypoxia.
Signal differences in RNA Pol II occupancy for promoters of active genes with a G4 (Pol II^{+} G4^{+}, black) and for promoters of active genes without a G4 (Pol II^{+} G4^{-}, blue) in cells treated with DMSO or pyPDS under hypoxic condition. Top panel, metagene plot of Pol II ChIP signal centered at the TSS showing the signal difference between DMSO- and pyPDS-treated hypoxic cells. Bottom panel, normalised Pol II ChIP-seq signal plotted for all individual loci and collectively represented by a heatmap plot. Increased signal is shown in blue, and decreased signal is shown from yellow to red. This shows that the retention of RNA Pol II after pyPDS treatment upon hypoxia is only observed at G4-marked (Pol II^{+} G4^{+}) promoters but not at non-G4-marked (Pol II^{+} G4^{-}) promoters, suggesting that it is the stabilisation of G4 structures by pyPDS causes Pol II retention.

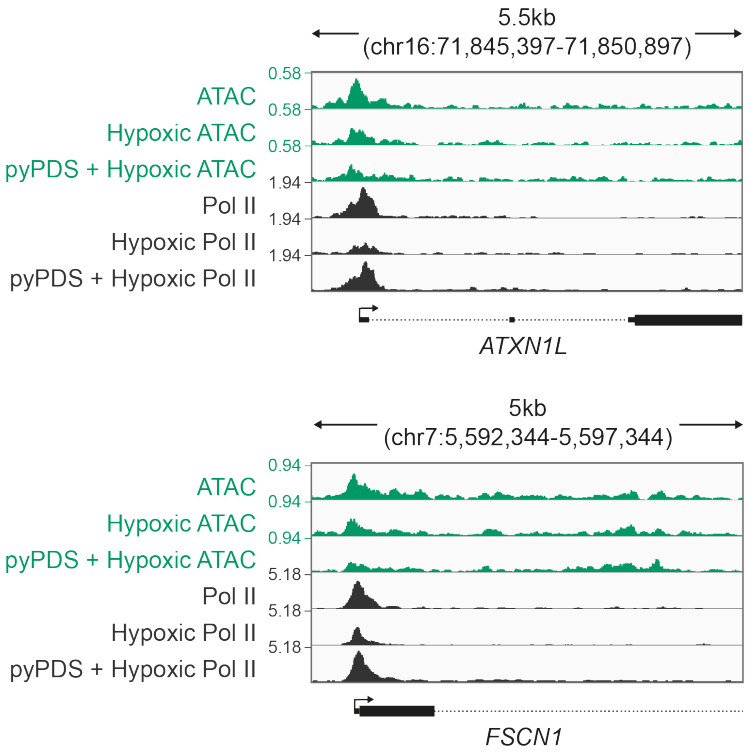

Figure 4.10: Stabilisation of G4s by pyPDS causes retention of RNA Pol II in spite of chromatin compaction upon hypoxia.

Genome browser view displaying examples of genes (e.g. *FSCN1* and *ATXN1L*) that show changes in chromatin accessibility (ATAC-seq, green) and RNA Pol II occupancy (Pol II ChIP-seq, black) for cells under normoxic (top), hypoxic (middle) and hypoxic conditions treated with pyPDS (bottom). Genomic coordinates are indicated above. This demonstrates that hypoxia results in reduced chromatin accessibility and RNA Pol II loading at promoters. An increase of RNA Pol II occupancy at the promoter is observed after pyPDS treatment, while the chromatin accessibility assessed by ATAC-seq remains reduced in pyPDS pre-treated hypoxic cells.

4.4 Summary

This chapter describes studies that set out to investigate the possible role for promoter G4s in RNA Pol II recruitment by chemically stabilising G4s during cellular hypoxia. To map the location of stabilised G4 sites, G4 ChIP-seq was performed following treatment of cells with G4 stabilising ligand pyPDS. However, an overall reduction of G4 ChIP-seq signal was observed after pyPDS stabilisation, including genes that were previously shown stabilised by G4 ligands *in vitro* (Siddiqui-Jain et al., 2002; Sun et al., 2011). Prior attempts at G4 ChIP-seq in the presence of G4 stabilising ligands within our laboratory have also shown similar decreases in G4 signal. As mentioned above, this may be due to DNA polymerase stalling at stable G4 structures, which is a well-established phenomenon (Han et al., 1999; Woodford et al., 1994). Polymerase stalling by ligand-stabilised G4s has also been widely used to characterise G4 formation (Chambers et al., 2015; Dexheimer et al., 2006; Qin et al., 2007). In G4 ChIP-seq, cells were cross-linked by formaldehyde, which reacts with the amino group of proteins or DNA to form a methylene bridge between molecules in close proximity (Hoffman et al., 2015). Furthermore, immunoprecipitated DNA libraries are amplified by polymerase chain reaction (PCR) and subjected to sequencing. Strand extension by DNA polymerase is a key element for both of these steps. In this chapter, I performed G4 ChIP-seq in the presence of pyPDS. The amino group in pyPDS may enable its reaction with formaldehyde and subsequent cross-linking to DNA. Such cross-linked pyPDS-stabilised G4s can arrest DNA polymerase progress during PCR or Illumina sequencing, reducing the amplification efficiency and thus result in reduced enrichment when compared to other non-G4-forming regions or background. This would result in reduced G4 ChIP signal at G4 peaks genome-wide as seen in my G4 ChIP-seq data. Interference with polymerase extension is more likely to impact amplification and sequencing of folded G4s and therefore have a greater influence on G4 ChIP-seq as opposed to other methods where G4s may be unfolded. In agreement, the G4-ligand mediated signal depletion was not observed in ATAC-seq and RNA Pol II ChIP-seq since these methods do not specifically enrich for folded G4s.

Furthermore, BG4 IF staining, a DNA polymerase-independent method, detected increase in G4 formation upon pyPDS treatment. BG4 IF is unable to specifically visualise G4 formation at promoters and instead can only provide a surrogate measurement of G4 folding. Despite this limitation, increase in the number of BG4 foci confirmed the increase in G4s upon

pyPDS treatment of normoxic cells. While depletion of oxygen substantially reduced BG4 foci number and IF signal intensity in cell nuclei, fewer G4s were lost upon pre-treatment with pyPDS, supporting my contention that prior stabilisation can protect G4s from unfolding in hypoxia.

Although pre-treatment of pyPDS retained G4s from unfolding in hypoxia, a reduction in chromatin accessibility was still observed in pyPDS-treated hypoxic cells at G4 marked active promoters. These results show that global chromatin compaction still ensues despite increased G4 stability, suggesting that the presence of a folded G4 cannot interfere with a hypoxia-driven increase in nucleosome density at promoters. Surprisingly, this also suggests that G4s are able to form within more compact chromatin, however, this remains to be directly demonstrated at the local level.

Remarkably, treatment of cells by pyPDS and thus pre-stabilisation of G4s leads to a reduced loss of RNA Pol II at promoter G4s upon hypoxia. This suggests that ligand stabilised G4s are able to diminish the impact of chromatin compaction on Pol II occupancy. Furthermore, RNA Pol II recruitment in homeostatic normoxic cells was not altered by pyPDS, which rules out the possible confounding effects of pyPDS on RNA Pol II occupancy. Moreover, the retention of RNA Pol II through pyPDS-mediated G4 stabilisation was specific to G4-marked promoters as opposed non-G4 marked promoters, strongly suggesting a G4-specific mechanism.

In summary, my results in Chapter 2 and Chapter 3 have suggested that chromatin status and not transcriptional activity determines whether G4s fold at promoters. The work in this chapter has further revealed that G4 formation can be modulated to augment RNA Pol II occupancy in an otherwise inhibitory chromatin environment. Overall, these results suggest that promoter G4 formation functions to enable transcription by promoting RNA Pol II recruitment.

Chapter 5

Discussion and conclusions

5.1 Promoter G4 folding, transcription activity and chromatin accessibility

Endogenous G4s have been found to predominantly fold within regulatory, nucleosome-depleted regions and mark sites of elevated transcription (Hänsel-Hertsch et al., 2016). While numerous other studies have indicated that G4s are associated with active gene transcription (Hou et al., 2019; Makowski et al., 2018; Raiber et al., 2012; Sengupta et al., 2019), there is some evidence to suggest that G4s at active gene promoters may form as a consequence of increased transcriptional activity (Duquette et al., 2004; Kouzine et al., 2004). However, these hypotheses are derived from *in vitro* systems or from genomic correlations of G4s, and the causal association between G4 formation and active transcription still remains to be tested in cells.

To investigate if promoter G4s form as a consequence of elevated transcription, I measured G4 folding using G4 ChIP-seq following chemical treatments to inhibit two different steps of Pol II transcription, i.e. DRB to inhibit elongation and TPL to inhibit initiation. DRB treatment abolishes ssDNA formation within the gene body but increases ssDNA formation at promoter regions (Wu et al., 2020), where the latter is likely due to the increased pausing of RNA Pol II near the TSS (Core et al., 2008). Surprisingly, no increase in G4 formation at promoters was detected following DRB treatment, implying that the increase in ssDNA formation does not result in increased promoter G4 folding. More importantly, no significant decrease in G4 signal intensity was observed at promoters, indicating that G4 formation does not rely on transcription elongation.

During transcription initiation, the local melting of promoter DNA is important to the formation of transcription bubble and the exposure of single-stranded template, which allows polymerase scanning (Cheung et al., 2011). The melting of promoter DNA is facilitated by negative supercoiling, which has been shown to induce the formation of G4s (Kouzine et al., 2004; Revyakin et al., 2004). It therefore still remained possible that

promoter G4s were established during transcription initiation. Inhibition of transcription initiation by TPL results in elimination of most ssDNA from transcribed regions and promoters (Wu et al., 2020). However, this did not cause any decrease in promoter G4 folding. These findings refute the hypothesis that increased ssDNA formation during the transcription process augments promoter G4 formation, and indicate that the folding of promoter G4s is not a consequence of transcriptional activity. Contrarily, TPL treatment resulted in increased G4 formation at many gene promoters. Since TPL functions by inhibition of Pol II-associated helicases XPB and XPD (Titov et al., 2011), this may result in reduced G4 unfolding and cause the observed increase in promoter G4s. However, the impact of alterations in cellular XPB/XPD activity on the G4 landscape has not yet been directly tested. G4 ChIP-seq performed in XPD-deficient cells (Armelini et al., 2005; Saffi et al., 2010) or upon siRNA-mediated depletion of XPB/XPD complex can provide further evidence to the role of these helicases in resolving promoter G4s.

Since R-loop formation also relies on transcriptional activity, the findings of my studies also suggest the folding of promoter G4s is independent of R-loops. This is not entirely surprising, as R-loops form downstream of the TSS (Chen et al., 2017), whereas the G4s are found primarily enriched upstream of the TSS (Hänsel-Hertsch et al., 2016). But to demonstrate this relationship directly, R-loop formation can be inhibited or augmented by overexpression of RNase H1 (De Magis et al., 2019) or catalytically dead RNase H1 (Chen et al., 2017) and its impact on promoter G4s can then be studied using G4 ChIP-seq. If my contention that R-loop formation does not promote G4 folding is valid, inhibition of R-loop formation will not result in the loss of G4 ChIP-seq signal.

My findings also raised the question that if not transcriptional activity then what determined the formation of G4 within gene promoters. G4 motifs have been found outside of nucleosome-bound regions within the human genome (Wong and Huppert, 2009). Corroborating this, endogenous G4s have been found predominantly in accessible chromatin regions (Hänsel-Hertsch et al., 2016). Moreover, G4 structures have been suggested to influence the positioning and occupancy of nucleosomes within chromatin (Kouzine et al., 2017). Increased chromatin accessibility following HDAC inhibition by entinostat was previously shown to cause elevated G4 formation at promoters. It can

therefore be hypothesised that chromatin environment defines where G4s folds, with promoters providing the most permissive environment. To directly test if G4 folding required highly accessible chromatin within promoters, I utilised the cellular response to acute hypoxia as a tool for inducing rapid chromatin compaction at promoters (Chen et al., 2006; Dutta et al., 2014; Gao et al., 2016; Kirmes et al., 2015). Greater than 85% of G4-marked promoters showed significantly decreased accessibility upon hypoxia as measured by ATAC-seq, which resulted in a significant loss of G4 ChIP signal from a fifth of promoter G4s. This clearly indicated that some promoter G4s are sensitive to chromatin condensation, however, majority of promoter G4s did not show a statistically significant decrease in G4 ChIP signal, despite showing chromatin compaction. Since G4 ChIP-seq demonstrates lower signal-to-noise ratio when compared to other ChIP-seq approaches (Hänsel-Hertsch et al., 2018), it may not be able to detect small changes in G4 formation as statistically significant. However, these findings could also indicate that some promoter G4s are resistant to chromatin compaction conditions under acute hypoxia. In support of this, G4-stabilising ligands may protect promoter G4s from unfolding as a result of compact chromatin (discussed below). To examine this, a precise mapping of nucleosome positions is needed. This could provide direct demonstration whether those promoter G4s that remain folded in hypoxia induced chromatin compaction are still located in nucleosome-free regions after hypoxia treatment or not.

In this study, chromatin accessibility is determined using ATAC-seq, which uses the hyperactive transposase Tn5 to define sites of accessible chromatin. But precise determination of nucleosome positioning using this methodology requires sequencing at great depth and the data in this study lack such detail. Therefore, it is still possible that despite showing a statistically significant decrease in ATAC-seq signal, DNA that comprises the G4 at the majority of gene promoters remains free of nucleosomes. This could provide an alternative explanation for why majority of promoter G4s are not lost despite chromatin compaction. To accurately assess nucleosome positioning, more sophisticated methods need to be employed, for example chemical cleavage of the nucleosome-dyad that enables the precise mapping of single nucleosomes and linkers *in vivo* (Chereji et al., 2018).

The mechanism of how promoter G4s become unfolded upon chromatin compaction in hypoxia is still unclear. Chronic hypoxia has previously been demonstrated to cause a reduction in APE1 protein levels (Chan et al., 2014). Abrogated activity of APE1, an endonuclease involved in base excision repair, could result in decreased G4 formation in cells (Roychoudhury et al., 2020). However, no changes in APE1 levels or activity were observed in my experiments, which suggests that G4 loss during acute hypoxia is through an alternative mechanism.

Chromatin compaction introduces positive torsion to facilitate nucleosome assembly (Kaczmarczyk et al., 2020). As positive torsion is thought to disfavour the G4 formation, it is possible that nucleosome assembly in compacted chromatin resolves the G4 structures via torsional stress. In addition, the loss of G4s may also be due to recruitment of G4-resolving helicases upon chromatin compaction. For example, hypoxia is associated with an increase in H3K9me3 (Chen et al., 2006), which recruits the G4-unfolding helicase ATRX to chromatin (Iwase et al., 2011; Law et al., 2010), which may in turn unwind G4s. To address this, ATRX ChIP-seq could be performed to see if ATRX is specifically recruited to sites which show G4 loss in hypoxia.

In hypoxia, the loss of promoter G4s was also accompanied with a corresponding loss of RNA Pol II occupancy. Since, the reduction in transcriptional activity does not lead to a loss of promoter G4s as indicated by the transcriptional inhibition experiments, it can be deduced that chromatin status is the primary determinant of promoter G4 folding in cells.

5.2 Role for promoter G4s in transcription

The role of promoter G4s in regulation of transcription is not yet understood. Since loss of promoter G4s was accompanied by a corresponding loss of RNA Pol II occupancy upon chromatin compaction in hypoxia, this suggests an association between folding of G4s at gene promoters and RNA Pol II occupancy. Since the formation of promoter G4s precedes transcription initiation, it can be hypothesised that loss of promoter G4s occurs prior to and may result in the loss of RNA Pol II.

Chemical biology approaches provide powerful tools for manipulating the stability of cellular G4s. Pre-treatment of cells with pyPDS, one such G4 stabilising ligand, could diminish the loss of promoter G4s upon hypoxia, thus counteracting the effects of chromatin compaction. Ligand-bound G4 demonstrate increased thermodynamic stability and thus may be resolved at a reduced rate. This has been shown for helicase DHX36, of which G4 unwinding activity is inhibited by G4-stabilising ligand, with the inhibition is attributed to thermal stability (Chen et al., 2015c). The reduced unwinding rates of helicase in the presence of G4 ligands would in turn allow increased retention of G4s upon chromatin compaction. Interestingly, despite the increase in cellular G4 content, hypoxia-associated chromatin compaction still ensued, implying that stabilisation by pyPDS could enable G4s to form within compact chromatin. However, I am unable to directly detect the folding of G4s at specific genomic locations in the presence of pyPDS due its incompatibility with G4 ChIP-seq (as discussed in Chapter 4; see section 4.4), which would allow for the direct demonstration of whether stabilised G4s fold in inaccessible chromatin.

Protecting the promoter G4 content with pyPDS also diminished the loss of RNA Pol II at the majority of sites that otherwise showed RNA Pol II loss upon hypoxia. Importantly, pyPDS did not impact the Pol II occupancy at non-G4 containing promoters. These findings directly demonstrate that stabilised G4s result in polymerase retention in otherwise non-conducive compacted chromatin. However, it still remains unknown whether this polymerase actively engages in transcription. The transcriptional output at such sites of Pol II retention could be addressed with nascent RNA sequencing following pyPDS treatment in hypoxia and comparing the transcriptional output with DMSO-treated normoxic and hypoxic cells.

The mechanism by which promoter G4s aid the recruitment of RNA Pol II is not yet understood. A recent study used a MYC promoter G4 oligonucleotide to bait interacting proteins from nuclear extracts and found subunits of RNA Pol II by mass spectrometry (Makowski et al., 2018). This suggests that promoter G4s may directly recruit Pol II via interactions with polymerase subunits, however direct demonstration of such an interaction is still lacking. Binding affinity measurements, which examine the direct interaction between G4 and reconstituted Pol II complex *in vitro*, could help address this question. Furthermore, G4s are likely to recruit RNA Pol II by providing high affinity interactions with TFs.

Supporting this, various studies have shown the interactions between G4s and TFs. For example, TFs such as SP1, XPB, CNBP and LARK (Borgognone et al., 2010; Gray et al., 2014; Niu et al., 2019; Raiber et al., 2012), bind to G4-forming oligonucleotides with binding affinity comparable to their defined double-stranded consensus binding motifs. Moreover, recent data indicate that G4s may act as TF binding hotspots that promote elevated transcription by providing binding sites for multiple TFs (Spiegel et al., 2021). To test whether TFs are involved in G4-mediated retention of Pol II in hypoxia, ChIP-seq experiments looking at occupancy of TFs known to bind G4s will need to be performed. These will help address the question whether the occupancy of TFs is also diminished due to chromatin compaction upon hypoxia and if pyPDS-mediated stabilisation would also rescue TF occupancy at sites where Pol II occupancy rescued.

Repressive chromatin is marked by histone modifications including H3K9me3, which drives formation of constitutive heterochromatin via recruitment of the heterochromatin protein (HP) family of proteins (Ninova et al., 2019). On the other hand, modifications such as H3K4me3 marks accessible chromatin containing sites of the RNA Pol II binding and active transcription (Barski et al., 2007). In specific biological contexts, such as in embryonic stem cells, promoters of genes poised for active transcription are often found to be bivalently marked, i.e. present repressive and activating histone modifications (Voigt et al., 2013). These genes are often transcriptionally silent but show polymerase loaded promoters. Since pre-stabilisation of G4s in hypoxia generates RNA Pol II loaded promoters within compacted chromatin, it is possible that these sites get marked by both active and inactive marks and obtain bivalency. To address this question, ChIP-seq for H3K9me3 and H3K4me3 would examine whether both marks co-exist at retained Pol II sites.

A recent discovery from our lab characterised G4 landscape in 22 different patient-derived tumour xenograft models and showed that promoter G4s are associated with highly amplified genes that show high expression (Hänsel-Hertsch et al., 2020). The findings in my thesis provide a possible mechanism that the increased occurrence of G4s at such amplified regions in itself may be sufficient in driving elevated transcriptional output in breast cancer, thus highlighting G4s as a potential target for future therapeutics approaches.

Hypoxia is a common phenomenon in the majority of solid tumours, and hypoxia-driven cellular response plays an important role in disease progression and resistance to therapy (Muz et al., 2015). Hypoxia also promotes a more malignant and invasive phenotype leading to poor prognosis (Semenza, 2010). Therapeutic approaches that interfere with the hypoxic response are therefore of great interest. Cellular response to hypoxia is achieved by several mechanisms including transcriptional regulations and energy conservation processes (Batie et al., 2018). Since it has been shown that pre-stabilisation of G4s in hypoxia augments the RNA Pol II loading, it would be interesting to explore if perturbing the cellular G4 landscape, in particular pre-stabilisation of G4s, could affect the hypoxia response in cells by altering the transcriptional profile. Moreover, hypoxia is known to augment cellular migration. It would be interesting to explore the impact of pre-treatment with G4 ligands on hypoxic cell migration and invasion using assays such as transwell migration assay or scratch-wound assay. In general, a better understanding on the role of G4s in cancer progression could further promote G4s as therapeutic targets in cancer.

5.3 Conclusions

The overall aim of this thesis was to investigate the causal relationship between G4 formation, active transcription and chromatin. The key outcomes are:

- Transcriptional inhibition does not alter the formation of G4s at gene promoters, indicating that promoter G4s precede active transcription.
- Promoter G4 folding is sensitive to chromatin compaction, suggesting that chromatin status is a determinant of G4 folding in cells.
- Stabilisation of G4s diminishes the loss of RNA Pol II at promoters in compact chromatin, demonstrating that G4s directly promote polymerase recruitment.

These results suggest a model for the relationship between promoter G4 folding, transcriptional activity and chromatin. During active transcription, G4s fold in accessible promoter regions and promote RNA Pol II recruitment (Figure 5.1A). Inhibiting transcription by DRB or TPL eventually causes the degradation of RNA Pol II, however, G4 formation is not perturbed (Figure 5.1B). By contrast, chromatin compaction causes G4 unfolding at some sites accompanied by the corresponding loss of RNA Pol II binding (Figure 5.1C). Ligand-

mediated G4 stabilisation (red star) can preserve promoter G4 folding under conditions of chromatin compaction, which in turn causes retention of RNA Pol II (Figure 5.1D).

In conclusion, these findings build on earlier work showing that presence of G4s at promoters correlates with increased gene expression and provide evidence to suggest that the G4 structure itself may be sufficient to drive transcription by promoting polymerase occupancy.

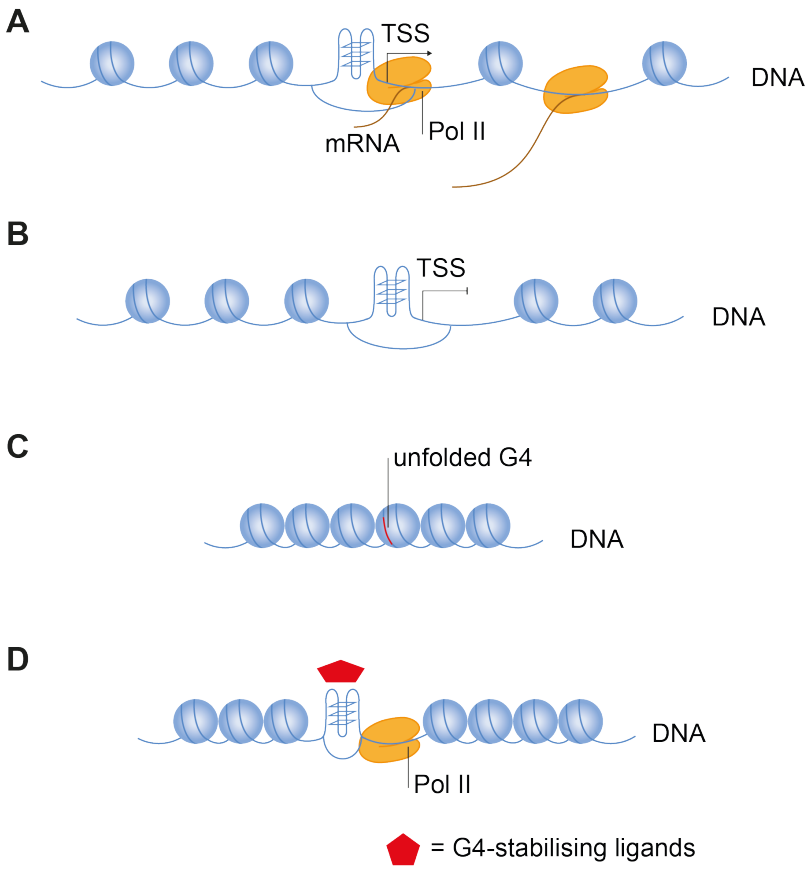

Figure 5.1: Model showing interactions between promoter G4s, RNA Pol II and chromatin status.
(A) The folding of promoter G4 is seen in accessible promoter regions. RNA Pol II is recruited to the G4 to promote transcription. **(B)** Transcriptional inhibition does not cause loss of promoter G4 folding. **(C)** Chromatin compaction causes G4 unfolding and loss of RNA Pol II binding. **(D)** Ligand-mediated G4 stabilisation (red star) can preserve promoter G4 folding under conditions of chromatin compaction, which in turn causes retention of RNA Pol II.

Chapter 6

Materials and methods

6.1 Mammalian cell culture

6.1.1 Cell culture and cryopreservation

K562 cells (ATCC, CCL-243) were cultured in RPMI-1640 (Gibco, 21875034) supplemented with 10% fetal bovine serum (FBS; Gibco, A3840402). U2OS cells (ATCC, HTB-96) were cultured in DMEM (Gibco, 41966029) supplemented with 10% FBS. Cells were cultured at 37 °C in 5% CO_2, 21% O_2 unless stated otherwise. Cell line genotypes were certified by the supplier. Cells lines were confirmed mycoplasma-free by RNA capture ELISA.

Cells were periodically cultured to between 80 and 90% confluence. For passaging, K562 cells were centrifuged for 3 min at 1,000 x *g*. Pelleted cells were then resuspended in fresh media and replated. U2OS cells were rinsed twice with sterile phosphate-buffered saline (PBS; 137 mM NaCl, 2.7 mM KCl, 4.3 mM Na_2HPO_4, 1.4 mM KH_2PO_4) and trypsinised using 0.05% trypsin-EDTA (Gibco, 25300054). DMEM containing 10% FBS was then added to neutralise the trypsinisation process, and cells were centrifuged at 1,000 x *g* for 3 min. The pelleted cells were resuspended in growth medium and replated into flasks. Cell seeding densities varied depending on the initial confluence.

To prepare frozen stocks, cells were resuspended in 90% FBS and 10% DMSO. The resuspension was aliquoted into cryogenic vials and frozen at -80 °C using a Mr. Frosty™ Freezing Container (Nalgene). Cells were then stored in liquid nitrogen for long-term preservation. To revive frozen cell lines, stocks were thawed rapidly at 37 °C and immediately diluted with growth medium. Cells were then centrifuged at 1,000 x *g* for 3 min before being resuspended in fresh growth medium and plated in flasks.

6.1.2 Hypoxia treatment

Cells were washed twice with PBS, resuspended in culture medium and placed in a hypoxia incubator with 1% O_2, 5% CO_2 and 94% N_2 at 37 °C for 1 h. Cell culture media was pre-equilibrated to 1% O_2 prior to use. Untreated cells, which were maintained at 37 °C with 5% CO_2, 21% O_2 until further usage, served as normoxic controls.

6.1.3 Drug treatments

Before drug treatments, cells were counted using Muse Cell Analyzer (Millipore) and seeded into flasks or plates. 5,6-dichlorobenzimidazole 1-ß-D-ribofuranoside (DRB; Sigma, D1916), triptolide (Sigma, T3652), garcinol (Santa Cruz, sc-200891) and pyPDS (Müller et al., 2010) were dissolved in DMSO, and diluted in growth medium to the following final concentrations. For DRB inhibition, cells were incubated with 100 µM DRB for 1 h. For triptolide treatment, cells were treated either with 125 nM triptolide for 1 h or with 10 µM triptolide for 30 min or 2 h. To perturb chromatin structure, cells were incubated with 5 µM garcinol for 4 h. For G4 stabilisation in hypoxia, cells were pre-treated with 1 µM pyPDS for 1 h before exposure to 1% O_2 in the presence of pyPDS for 1 h. Cells pre-treated with 1 µM pyPDS for 1 h and maintained in normoxic conditions served as controls.

6.2 Western blotting

6.2.1 Preparation of whole cell lysates

To avoid protein denaturation and degradation, lysates were prepared on ice. 5 million cells were centrifuged 1,000 x *g* for 3 min, media was discarded, and cells were washed twice with ice-cold PBS. Cells were lysed in Pierce® RIPA lysis buffer (Thermo Scientific, 89900), supplemented with phosSTOP (Sigma, 4906845001) phosphatase inhibitor and cOmplete Protease Inhibitor Cocktail (Roche, 11697498001), by sonication using a Bioruptor Plus (30 seconds on/60 seconds off for 5 cycles; Diagenode). The lysate was centrifuged at maximum speed for 5 min to remove cell debris and quantified using a Direct Detect Spectrometer (Millipore).

6.2.2 Sodium dodecyl sulphate (SDS)-polyacrylamide gel electrophoresis (PAGE)

Cell lysates containing 10-20 µg of protein per lane were resolved by NuPAGE 4-12% Bis-Tris gels or NuPAGE 3-8% Tris-Acetate gels (Invitrogen) depending on the size of target proteins. Precision Plus Protein Dual Color Standards (Biorad, 1610374) were used as molecular weight markers to determine protein size. Prior to loading, samples were boiled at 70 ˚C for 10 min in 1× NuPAGE LDS sample buffer (Invitrogen, NP0007) and 1× NuPAGE sample reducing agent (Invitrogen, NP0009). The gel tank apparatus was filled with 1× MOPS (3-(N-morpholino) propanesulfonic acid) SDS running buffer (Invitrogen) or 1× Tris-Acetate SDS running buffer (Invitrogen) depending on the gel used. Electrophoresis was carried out at 60 V for the first 30 minutes and then at 120 V for varying durations depending on target protein size.

6.2.3 Immunoblotting

After SDS-PAGE, separated proteins were transferred onto a nitrocellulose membrane using the iBlot® 2 Dry Transfer System (Invitrogen) or wet transfer (for RNA Pol II and phosphorylated Pol II). Wet transfer was carried out in 1× NuPAGE transfer buffer (Invitrogen) with 20% methanol at 100 V for 1.5 h at room temperature (RT) or 30 V overnight at 4 ˚C. For overnight transfers, an ice pack was placed in the tank to prevent heating of the system. Subsequently, the membrane was washed with tris-buffered saline (TBS; 50 mM Tris-HCl with pH 7.5, 150 mM NaCl) and blocked using Odyssey Blocking Buffer TBS (Li-Cor) for 1 h at RT. The membrane was then incubated with primary antibody diluted in Odyssey Blocking Buffer TBS containing 0.1% Tween-20 (VWR chemicals). All primary antibodies (Table 6.1) were used at a concentration of 1:1,000, apart from the HIF1α antibody which was used at a concentration of 1:500. After incubation with primary antibodies, the membranes were washed three times in TBS containing 0.1% Tween-20 (TBS-T) for 10 min each. Following incubation with fluorescent secondary antibodies (LiCor, IRDye) for 45 min at RT, the membranes were washed with TBS-T three times and with TBS one more time. The membranes were visualised using Odyssey CLx Imaging System (LiCor) and Image Studio Version 4.0 software. Protein bands were quantified using ImageJ.

Table 6.1: Antibodies used for Western blotting in human cells.
The supplier and the catalogue numbers are listed for the primary antibodies used for Western blotting.

Target protein	Supplier	Catalogue No.
APE1	Novus Biologicals	NB100-116
APE1 (Acetyl-Lys6)	Aviva Systems Biology	OAAF08210
ß-actin	Cell Signaling Technology	4970
	Sigma	A5441
HIF1α	BD Biosciences	610958
Histone H3 acetyl	Active Motif	39140
Histone H4 acetyl	Active Motif	39244
RNA Pol II	Abcam	ab817
RNA Pol II phospho Ser2	Active Motif	61984
RNA Pol II phospho Ser5	Abcam	ab5131

6.3 BG4 antibody purification

6.3.1 BG4 expression

BL21 (DE3) competent *E.coli* cells (New England Biolabs, C2527H) were transformed with the BG4 expression plasmid pSANG10-3F-BG4 according to the manufacturer's instructions. *E.coli* cells carrying the plasmids were inoculated in 2 ml 2×TY medium supplemented with 50 µg/ml kanamycin and 2% glucose, and were grown overnight at 30 ˚C shaking at 200 rpm. 200 µl of *E.coli* cell culture was then inoculated into 100 ml auto-induction media (see recipes below). Cultures were grown at 37 ˚C with shaking at 250 rpm for 6 h, before overnight incubation at 25 ˚C with shaking at 280 rpm.

Auto-induction medium

2 mM $MgSO_4$

0.2 × 1000 × metals mix

1 × 5052 buffer

1 × M buffer

Dissolved in basic ZY medium, autoclaved and stored at RT.

1000 × metals mix

50 mM	$FeCl_3 \bullet 6H_2O$ (dissolved in 0.1 M HCl)
20 mM	$CaCl_2$
10 mM	$MnCl_2 \bullet 4H_2O$
10 mM	$ZnSO_4 \bullet 7H_2O$
2 mM	$CoCl_2 \bullet 6H_2O$
2 mM	$CuCl_2 \bullet 2H_2O$
2 mM	$NiCl_2 \bullet 6H_2O$
2 mM	$Na_2MoO_4 \bullet 2H_2O$
2 mM	$Na_2SeO_3 \bullet 5H_2O$
2 mM	H_3BO_3

Individual stock solutions (except the $FeCl_3$ in HCl) were autoclaved, mixed and stored at RT.

50 × 5052 buffer

25% (v/v)	glycerol
2.5% (w/v)	glucose
10% (w/v)	α-lactose

Dissolved in H_2O, autoclaved and stored at RT.

50 × M buffer

1.25 M	KH_2PO_4
2.5 M	NH_4Cl
0.25 M	Na_2SO_4

Dissolved in H_2O, autoclaved and stored at RT.

ZY medium

10 g	N-Z amine A (Sigma, N4517)
5 g	yeast extracts (Sigma, 70161)

Dissolved in 1 L H_2O, autoclaved and stored at RT.

6.3.2 BG4 purification

E.coli cells were collected by centrifugation at 4,000 x *g*, 4 ˚C for 30 min. The cell pellets were resuspended in 8 ml ice-cold TES buffer (50 mM Tris-HCl pH 8.0, 1 mM EDTA pH 8.0, 20% sucrose) and incubated on ice for 10 min. The suspension was incubated on ice in 12 ml ice-cold TES diluted 1:5, with benzonase and 2 mM $MgSO_4$, for another 15 min and then centrifuged for 10 min at 20,000 x *g* at 4 ˚C. Supernatants were filtered before being incubated (rotating for 1 h at RT) with PBS-washed His-select Nickel affinity beads (Sigma, P6611). The bead samples were loaded into the Proteus 1-step batch Midi Spin columns (Generon, GEN-1SB08), rinsed twice with ice-cold washing buffer (100 mM NaCl, 10 mM imidazole in PBS, pH 8.0) and eluted with elution buffer (250 mM imidazole in PBS, pH 8.0). The eluted fractions were then dialysed twice in PBS at 4 °C overnight using GeBaflex tubes (Generon, D045). Dialysed samples were aliquoted, snapfrozen in liquid nitrogen and stored at -80 ˚C.

6.3.3 Antibody analysis and quantification

Protein samples were loaded and resolved by SDS-PAGE as described in section 6.2.2, with a titration of bovine serum albumin (BSA) included as protein standards. Gels were then stained for 30 min with Instant Blue (Sigma, ISB1L) and rinsed with water 5 times. Gel images were visualised using the Odyssey CLx Imaging System (LiCor) and protein concentrations were quantified by comparison with BSA standards using ImageJ.

6.4 Enzyme-linked immunosorbent assay (ELISA)

6.4.1 Oligo annealing

Biotinylated oligonucleotides (Table 6.2; Sigma) were dissolved in 10 mM Tris-HCl, pH 7.4 to 100 µM and were annealed in 10 mM Tris-HCl (pH 7.4), 100 mM KCl at 5 µM by heating at 95 ˚C for 10 min and then slowly cooling down to 25 ˚C at a rate of 0.2 ˚C min^{-1}.

Table 6.2: Sequences of oligonucleotides used in ELISA.
Sequences of oligonucleotides used for assaying BG4 binding affinity. G-tracts that can form G4 structures in *MYC* G4 oligo are in bold and underlined. Mutations in the non-G4-forming control oligo (*MYC*-mut.) are highlighted in orange.

Name	Sequence
MYC G4	TGA**GGG**T**GGG**TA**GGG**T**GGG**TAA
MYC-mut.	TGAGCGTGCGTAGCGTGCGTAA

6.4.2 ELISA

Streptavidin-coated 96-well plates were hydrated in PBS for 30 min, incubated with 50 nM biotinylated oligonucleotides at RT for 1 h and washed with ELISA buffer (100 mM KCl, 50 mM KH_2PO_4 adjusted to pH 7.4 using KOH) three times. Plates were then blocked for 1 h in blocking buffer at RT. After blocking, plates were incubated with serial dilutions of BG4 (from 200 nM to 0 nM diluted in blocking buffer) for 1 h, washed with ELISA buffer/0.1% Tween 20 three times and incubated with HRP-conjugated anti-FLAG antibody (diluted 1:15,000 in blocking buffer; Sigma, A8592) for 1 h. Following three washes with ELISA buffer/0.1% Tween 20, plates were dried and 3,3',5,5'-tetramethylbenzidine (TMB; Sigma, T4319) was added. The reaction was stopped with 2 M HCl. Signal intensity was measured at 450 nm using a PHERAstar microplate reader (BMG Labtech). Dissociation constants were calculated from binding curves using GraphPad Prism (GraphPad Software Inc.) and the standard error of the mean (s.e.m.) was calculated from three replicates. The following equation was used: $y = (B_{max}*x)/(K_d + x)$, where B_{max} is the maximum specific binding (i.e. saturation) and x is the concentration of BG4 (nM).

6.5 Micrococcal nuclease (MNase) digestion

1 million cells were washed in ice-cold PBS and incubated in lysis buffer (10 mM Tris pH 8, 10 mM $MgCl_2$, 0.5% NP-40, fresh PIC + 1mM DTT) at 4 °C for 10 min. Nuclei were pelleted by centrifugation at 1,400 x *g* for 5 min at 4°C and washed once with lysis buffer. Samples were digested with 0.5 U MNase at 25 °C in digestion buffer (15 mM Tris pH 7.5, 60 mM KCl, 15 mM NaCl, 250 mM Sucrose, 1 mM $CaCl_2$, fresh PIC + 1mM DTT) and the reaction stopped with an equal volume MNase stop solution (40 mM EDTA + 0.5% SDS). Fragmented DNA was incubated with RNase and Proteinase K overnight at 42 ˚C and purified using a QIAGEN

MinElute kit. Equal amounts of DNA (300 ng) from each sample were then loaded and resolved on 2% E-Gel EX precast agarose gels (Thermo Fisher, G800802).

6.6 Assay for transposase-accessible chromatin using sequencing (ATAC-seq)

ATAC-seq was performed with three biological replicates as previously described (Calviello et al., 2019). Briefly, 50,000 K562 cells were collected and incubated with transposase Tn5 (Illumina, 20034198) at 37 °C. After 1-hour incubation, tagmented DNA samples were amplified using the Nextera Index kit (Illumina, FC-121-1030). DNA fractions were size selected (100-1,000 bp) using AMPure XP beads (Beckman Coulter, A63880) according to the manufacturer's instructions. Libraries were sequenced in paired-end mode with 75-bp read length on the NextSeq 500 platform (Illumina).

6.7 G4 mapping by chromatin immunoprecipitation followed by high-throughput sequencing (G4 ChIP-seq)

G4 ChIP-seq in K562 cells was performed using the G4-specific antibody BG4 essentially as described previously (Hänsel-Hertsch et al., 2018). In brief, 25 million K562 cells were fixed in fresh culture medium containing 1% formaldehyde (Thermo Fisher, 28908) for 10 min at RT followed by quenching with glycine to a final concentration of 125 mM for 10 min. The cell pellet was then wahsed twice with ice-cold PBS and chromatin was isolated using lysis and hypotonic buffer (Chromatrap, 100008) following the manufacturer's instructions. Chromatin was sonicated into 100-500bp fragments using a Bioruptor Plus (Diagnode) at 4 °C and quantified via Qubit assays (Thermo Fisher).

At least three biological replicates were performed. For each biological replicate, three independent technical replicates and matched inputs were performed. About 300 ng of chromatin was added in each reaction. For immunoprecipitation, 0.3 µg of G4 structure-specific antibody BG4 was added except for the input sample. Following washing steps, immunoprecipitates and inputs were treated in TE buffer supplemented with Proteinase K (Invitrogen, AM2546). The reaction was then purified using a QIAGEN MinElute Kit (QIAGEN, 28206). For library preparation, up to 5 ng of the DNA was added in each reaction and the reaction was performed using Nextera library preparation kit (Illumina, 20034197) following

the manufacturer's instructions with some modifications. For a 40 µl reaction, 1.25 µl of Tn5 enzyme was used. The library was amplified using Nextera index primers (Illumina, FC-121-1030) and purified using QIAGEN MinElute kit. Samples were then subjected to single-end sequencing with a read length of 75 bp on an Illumina NextSeq 500 instrument.

6.8 RNA polymerase II (RNA Pol II) ChIP-seq

RNA Pol II ChIP-seq was performed essentially as previously described, with five biological replicates (Varshney et al., 2015). About 10 million K562 cells were fixed using 1% formaldehyde for 10 min at RT and the reaction was quenched with 125 mM glycine for another 10 min. Cells were rinsed with ice-cold PBS twice and with PBS/0.5% NP-40 twice. Subsequently, cells were incubated in high salt buffer (1 M NaCl, 0.5% NP-40 in PBS) for 30 min on ice, followed by incubation with low salt buffer (0.1 M NaCl, 10 mM Tris pH 8, 1 mM EDTA and 0.1% NP-40) for 30 min on ice. Cells were then cleaned by centrifugation through a 100 mM sucrose cushion (100 mM sucrose in low salt buffer) and the chromatin was sonicated into 100-500 bp fragments using a Bioruptor Plus (Diagnode) at 4 °C. Qubit assays (Thermo Fisher) were performed for chromatin quantification.

0.5 µg of DNA was put aside as input and 15 µg of chromatin was immunoprecipitated overnight with 5 µg RNA Pol II antibody (abcam, ab817) bound to protein A/G sepharose (Thermo Fisher, 10002D). Immunoprecipitates were washed three times in RIPA buffer (150 mM NaCl, 50 mM Tris pH 8, 1% IGEPAL CA-630, 0.1% SDS and 0.5% sodium deoxycholate), three times in 250 mM LiCl buffer (10 mM Tris pH 8, 250 mM LiCl, 1 mM EDTA, 0.5% IGEPAL CA-630 and 0.5% sodium deoxycholate) and three times in TE buffer. The immunoprecipitated DNA and input DNA were retrieved in elution buffer (1% SDS, 0.125mg/ml Proteinase K in TE) overnight at 42 °C and purified with QIAGEN MinElute kit according to the manufacturers' instructions. Libraries were prepared using NEBNext® Ultra™ II DNA Library Prep Kit for Illumina (NEB, E7645) and sequenced (75 bp read length in single-end mode) on an Illumina NextSeq 500 instrument.

6.9 Immunofluorescence (IF) staining

BG4 IF staining was performed essentially as described previously (Biffi et al., 2013b). U2OS cells grown on glass coverslips were fixed in PBS containing 4% (v/v) paraformaldehyde,

permeabilised with 0.1% (v/v) triton X-100 (Sigma, X100) in PBS and blocked with 2% (w/v) milk (Marvel, Premier Foods PLC) in PBS for 1 h. After blocking, cells were treated with 250 µg/ml RNaseA (Thermo Fisher, EN0531) at 37 ˚C for 1 h. Cells were then incubated with BG4 (1:800 diluted in PBS) followed by incubation with secondary rabbit anti-FLAG (Cell Signaling Technology, 2368) at 37 ˚C for 1 h, and subsequently incubated with tertiary Alexa Fluor 633 conjugated goat anti-rabbit (Thermo Fisher, A21071) antibody at 37 ˚C for 30 min. Coverslips incubated with 60 units of Turbo DNase (Invitrogen, AM2238) and coverslips incubated without BG4 antibodies provided negative controls. Following DAPI staining, coverslips were mounted with Prolong Gold (Thermo Fisher, P36961). Digital images were taken using a TCS SP5 confocal microscope (Leica) with Zeiss Zen software. BG4 foci number in nuclei were analysed with Icy software (De Chaumont et al., 2012) and BG4 signal density was quantified using ImageJ. 100–200 nuclei were counted per condition and graphs were plotted using GraphPad Prism (GraphPad Software Inc.).

6.10 Data analysis

Bioinformatic analyses were performed by Dr Angela Simeone, with input and discussion from Jiazhen Shen and Dr Dhaval Varshney.

6.10.1 Human reference genome and relative genomic annotation

Human genome assembly hg38 was downloaded from UCSC (hgdownload.cse.ucsc.edu/goldenPath/hg38/bigZips/hg38.fa.gz). Human annotations (gtf file) were downloaded from Genecode project portal (ftp.ebi.ac.uk/pub/databases/gencode/Gencode_human/release_28/gencode.v28.annotation.gtf.gz, Release 28 GRCh38.p12). Annotations for genomic regions (i.e. exons, introns, intergenic regions, 3’ UTR, 5’ UTR and 58381 promoters of all coding and non-coding genes defined as TSS ± 500bp) were extracted from the gtf file.

6.10.2 ATAC-seq data analysis

Fastq reads were trimmed from adapters using cutadapt (Martin, 2011) (cutadapt -a AGATCGGAAGAGC -A AGATCGGAAGAG). Resulting reads were aligned to hg38 with bwa mem -M -t 12. Bam files were generated by using samtools view -Sb -F780 -q 10 -L (ver: 1.8) (Li et al., 2009). All libraries were sequenced twice and processed and aligned separately.

Resulting alignments were merged and sorted. Duplicates were marked by Picard MarkDuplicates (ver: 2.20.3, http://broadinstitute.github.io/picard) and removed. Fragment size distribution was estimated using bam files containing uniquely mapped reads with Picard CollectInsertSizeMetric. To assess the amount of mitochondrial contamination, reads mapping to ChM were identified and counted directly from the alignment bam files. These mitochondrial reads were removed. For each library, regions with local accessibility were identified by calling peaks with macs2 with default options and excluding chrM. For each experimental condition, peak regions observed in 2 out of 3 biological replicates were selected as the consensus regions using bedtools multiIntersectBed (ver: 2.27.1) (Quinlan and Hall, 2010).

6.10.3 G4 ChIP-seq and RNA Pol II ChIP-seq data analysis

Fastq reads were trimmed from adapters using cutadapt (-m 10 -q 20 -O 3 -a CTGTCTCTTATACACATCT) and aligned to the human genome hg38 with bwa mem. Bam files were generated from alignment with samtools view and duplicated reads were marked and removed using Picard MarkDuplicates. The total number of unique reads was quantified for each library. Regions with local enrichments were obtained by calling peaks with macs2 for each individual pull-down library paired to the corresponding input control. For G4 ChIP-seq experiments, consensus regions of each biological replicate were defined as those observed in 2 out of 3 technical replicates (multiIntersectBed). The consensus for each experimental condition (across biological replicates) was obtained by selecting regions reproducibly observed in at least 2 of the 3 biological replicates. For Pol II ChIP-seq experiments, the consensus regions in each experimental condition were defined as the regions observed in at least 3 of the 5 biological replicates.

For all ChIP-seq experiments, genome-wide reads per million (RPM) were obtained by quantifying the read coverage across the genome and scaling it to a factor that reflected the individual library size (deeptools bamCoverage (Ramírez et al., 2014) - scaleFactor, where factor = 1,000,000/Lib_size). Similarity across individual libraries was evaluated based on RPM at consensus regions within each experimental condition. The use of G4 or RNA Pol II “signal” in this thesis refers to the median RPM for the specified experimental condition (across biological/technical replicates). G4 consensus regions were compared to the

accessible consensus regions (ATAC) observed in the same experimental conditions and quantified in terms of percentage of overlap; G4 consensus regions were compared to G4 motifs that can form G4s *in vitro* (G4-seq; Marsico et al., 2019) and quantified by evaluating the percentage that overlap.

6.10.4 Characterisation of G4 fold-enrichment at sites of interests

The genomic coordinates of various sites of interest were provided in bed format. The fold enrichment of these sites over random chance was computed using the Genomic Association Tester (GAT, https://gat.readthedocs.io/en/latest/contents.html, 1000 randomizations) and the analysis was restricted to the human whitelist. The overlap between G4 consensus sites (see section 6.10.3) and the segments to query against were identified, and their fold-enrichment over randomisations was calculated.

6.10.5 Density plots of genomic signals at transcription start sites (TSSs)

The metagene density signal profile at TSSs was produced in a similar manner for G4, RNA Pol II and ATAC-seq data. After creating bed files with set(s) of regions of interest, the deeptools function *computeMatrix* was performed on BW files containing RPM signals. Next, RPM signals were combined by averaging replicates of the same cell and then generating a new matrix of signals. The difference in normalised summarised signal between two experimental conditions was obtained by subtracting the matrix of normalised signal in one condition from the matrix of the other condition under investigation. Heatmaps were produced using deeptools plotHeatmap (options: --averageTypeSummaryPlot median). Metagene density signal profile plots were obtained by plotting the average trend of the signal of interest (i.e. the difference between two experimental conditions).

6.10.6 Differential signal analysis

All differential signal analyses were carried out with the R package edgeR. Initially, library size and read coverages at the regions of interest were computed. Prior to differential testing, the average cpm (counts per million) signal was estimated across all input libraries and a threshold value was defined as 2 times the 99^{th} quantile of the average distribution of input cpm. Subsequently, regions for which at least one pull-down library exceeded the threshold value previously defined were kept for subsequent analysis. This step was

performed for each sequencing assay independently and only cases where input libraries were available (G4 ChIP-seq, RNA Pol II ChIP-seq). A generalised linear model (*glmLRT*) with default parameters (negative binomial log-linear distribution of read counts) was used to assess regions with differential signal. For the differential test, batch information, biological and technical (when present) replicates were incorporated in the definition of the design matrix. The differential signal analysis compared pairs of experimental conditions: hypoxia vs normoxia (G4 ChIP-seq, ATAC-seq, RNA Pol II ChIP-seq); all pairwise comparison among: DMSO hypoxia, DMSO normoxia, pyPDS hypoxia and pyPDS normoxia. Regions with differential signal were identified as those with p-value < 0.05. The regions of interest used to test the effects of hypoxia plus pyPDS or hypoxia alone were defined as the G4 consensus regions overlapping Pol II consensus regions in normoxia at promoters (TSS ± 500bp). In the case of DRB and TPL experiments, the regions used for the differential signal analysis were obtained by merging the regions observed in the treatment and control case for each experimental condition respectively. To illustrate the outcome of the differential analysis, Bland Altman (MA) plots showing the average CPM (x-axis) versus the log-fold change (y-axis) obtained comparing the 2 conditions of interest were used.

References

Adey, A., Morrison, H.G., Asan, Xun, X., Kitzman, J.O., Turner, E.H., Stackhouse, B., MacKenzie, A.P., Caruccio, N.C., Zhang, X., et al. (2010). Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition. Genome Biol. *11*, R119.

Agalioti, T., Chen, G., and Thanos, D. (2002). Deciphering the transcriptional histone acetylation code for a human gene. Cell *111*, 381–392.

Agarwal, T., Roy, S., Kumar, S., Chakraborty, T.K., and Maiti, S. (2014). In the sense of transcription regulation by G-quadruplexes: Asymmetric effects in sense and antisense strands. Biochemistry *53*, 3711–3718.

Agrawal, P., Hatzakis, E., Guo, K., Carver, M., and Yang, D. (2013). Solution structure of the major G-quadruplex formed in the human VEGF promoter in K+: Insights into loop interactions of the parallel G-quadruplexes. Nucleic Acids Res. *41*, 10584–10592.

Amato, J., Madanayake, T.W., Iaccarino, N., Novellino, E., Randazzo, A., Hurley, L.H., and Pagano, B. (2018). HMGB1 binds to the KRAS promoter G-quadruplex: A new player in oncogene transcriptional regulation? Chem. Commun. *54*, 9442–9445.

Ambrus, A., Chen, D., Dai, J., Jones, R.A., and Yang, D. (2005). Solution structure of the biologically relevant G-quadruplex element in the human c-MYC promoter. Implications for G-quadruplex stabilization. Biochemistry *44*, 2048–2058.

Andersen, K.R., Leska, N.C., and Schwartz, T.U. (2013). Optimized E. coli expression strain LOBSTR eliminates common contaminants from His-tag purification. Proteins *81*, 1857–1861.

Di Antonio, M., Biffi, G., Mariani, A., Raiber, E.-A., Rodriguez, R., and Balasubramanian, S. (2012). Selective RNA Versus DNA G-Quadruplex Targeting by In Situ Click Chemistry. Angew. Chem. Int. Ed. *124*, 11235–11240.

Di Antonio, M., Ponjavic, A., Radzevičius, A., Ranasinghe, R.T., Catalano, M., Zhang, X., Shen, J., Needham, L.M., Lee, S.F., Klenerman, D., et al. (2020). Single-molecule visualization of DNA G-quadruplex formation in live cells. Nat. Chem. *12*, 832–837.

Armache, K.J., Kettenberger, H., and Cramer, P. (2003). Architecture of initiation-competent 12-subunit RNA polymerase II. Proc. Natl. Acad. Sci. U. S. A. *100*, 6964–6968.

Armelini, M.G., Muotri, A.R., Marchetto, M.C.N., De Lima-Bessa, K.M., Sarasin, A., and Menck, C.F.M. (2005). Restoring DNA repair capacity of cells from three distinct diseases by XPD gene-recombinant adenovirus. Cancer Gene Ther. *12*, 389–396.

Azzalin, C.M., Reichenbach, P., Khoriauli, L., Giulotto, E., and Lingner, J. (2007). Telomeric repeat-containing RNA and RNA surveillance factors at mammalian chromosome ends.

Science *318*, 798–801.

Balasubramanyam, K., Swaminathan, V., Ranganathan, A., and Kundu, T.K. (2003). Small molecule modulators of histone acetyltransferase p300. J. Biol. Chem. *278*, 19134–19140.

Balasubramanyam, K., Altaf, M., Varier, R.A., Swaminathan, V., Ravindran, A., Sadhale, P.P., and Kundu, T.K. (2004). Polyisoprenylated benzophenone, garcinol, a natural histone acetyltransferase inhibitor, represses chromatin transcription and alters global gene expression. J. Biol. Chem. *279*, 33716–33726.

Bang, I. (1910). Untersuchungen über die Guanylsäre. Biochem. Z. *26*, 293–311.

Bannister, A.J., and Kouzarides, T. (2011). Regulation of chromatin by histone modifications. Cell Res. *21*, 381–395.

Bao, H.L., and Xu, Y. (2020). Telomeric DNA-RNA-hybrid G-quadruplex exists in environmental conditions of HeLa cells. Chem. Commun. *56*, 6547–6550.

Bao, X., Rubin, A.J., Qu, K., Zhang, J., Giresi, P.G., Chang, H.Y., and Khavari, P.A. (2015). A novel ATAC-seq approach reveals lineage-specific reinforcement of the open chromatin landscape via cooperation between BAF and p63. Genome Biol. *16*, 1–17.

Barman, H.K., Takami, Y., Ono, T., Nishijima, H., Sanematsu, F., Shibahara, K. ichi, and Nakayama, T. (2006). Histone acetyltransferase 1 is dispensable for replication-coupled chromatin assembly but contributes to recover DNA damages created following replication blockage in vertebrate cells. Biochem. Biophys. Res. Commun. *345*, 1547–1557.

Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-Resolution Profiling of Histone Methylations in the Human Genome. Cell *129*, 823–837.

Batie, M., del Peso, L., and Rocha, S. (2018). Hypoxia and chromatin: A focus on transcriptional repression mechanisms. Biomedicines *6*, 1–19.

Baumann, P. (2010). G-Quadruplex DNA. Springer US *608*.

Baumli, S., Endicott, J.A., and Johnson, L.N. (2010). Halogen bonds form the basis for selective P-TEFb inhibition by DRB. Chem. Biol. *17*, 931–936.

Bedrat, A., Lacroix, L., and Mergny, J.L. (2016). Re-evaluation of G-quadruplex propensity with G4Hunter. Nucleic Acids Res. *44*, 1746–1759.

Bejugam, M., Sewitz, S., Shirude, P.S., Rodriguez, R., Shahid, R., and Balasubramanian, S. (2007). Trisubstituted isoalloxazines as a new class of G-quadruplex binding ligands: Small molecule regulation of c-kit oncogene expression. J. Am. Chem. Soc. *129*, 12926–12927.

Belmonte-Reche, E., and Morales, J.C. (2020). G4-iM Grinder: when size and frequency matter. G-Quadruplex, i-Motif and higher order structure search and analysis tool. NAR

Genomics Bioinforma. *2*, 1–12.

Belotserkovskii, B.P., Shin, J.H.S., and Hanawalt, P.C. (2017). Strong transcription blockage mediated by R-loop formation within a G-rich homopurine-homopyrimidine sequence localized in the vicinity of the promoter. Nucleic Acids Res. *45*, 6589–6599.

Benhalevy, D., Gupta, S.K., Danan, C.H., Ghosal, S., Sun, H.W., Kazemier, H.G., Paeschke, K., Hafner, M., and Juranek, S.A. (2017). The Human CCHC-type Zinc Finger Nucleic Acid-Binding Protein Binds G-Rich Elements in Target mRNA Coding Sequences and Promotes Translation. Cell Rep. *18*, 2979–2990.

Biffi, G., Tannahill, D., and Balasubramanian, S. (2012). An intramolecular G-quadruplex structure is required for binding of telomeric repeat-containing RNA to the telomeric protein TRF2. J. Am. Chem. Soc. *134*, 11974–11976.

Biffi, G., Di Antonio, M., Tannahill, D., and Balasubramanian, S. (2013a). Visualization and selective chemical targeting of RNA G-quadruplex structures in the cytoplasm of human cells. Nat. Chem. *6*, 75–80.

Biffi, G., Tannahill, D., McCafferty, J., and Balasubramanian, S. (2013b). Quantitative visualization of DNA G-quadruplex structures in human cells. Nat. Chem. *5*, 182–186.

Biffi, G., Tannahill, D., Miller, J., Howat, W.J., and Balasubramanian, S. (2014). Elevated levels of G-quadruplex formation in human stomach and liver cancer tissues. PLoS One *9*.

Bochman, M.L., Paeschke, K., and Zakian, V.A. (2012). DNA secondary structures : stability and function of G□quadruplex structures. Nat. Rev. Genet. *13*, 770–780.

Borgognone, M., Armas, P., and Calcaterra, N.B. (2010). Cellular nucleic-acid-binding protein, a transcriptional enhancer of c-Myc, promotes the formation of parallel G-quadruplexes. Biochem. J. *428*, 491–498.

Bourdoncle, A., Torres, A.E., Gosse, C., Lacroix, L., Vekhoff, P., Le Saux, T., Jullien, L., and Mergny, J.L. (2006). Quadruplex-based molecular beacons as tunable DNA probes. J. Am. Chem. Soc. *128*, 11094–11105.

Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat. Methods *10*, 1213–1218.

Buenrostro, J.D., Wu, B., Chang, H.Y., and Greenleaf, W.J. (2015). ATAC-seq : A Method for Assaying Chromatin Accessibility Genome-Wide. Curr. Protoc. Mol. Biol. 1–9.

Bugaut, A., and Balasubramanian, S. (2008). A Sequence-Independent Study of the Influence of Short Loop Lengths on the Stability and Topology of Intramolecular DNA G-Quadruplexes. Biochemistry *47*, 689–697.

Burge, S., Parkinson, G.N., Hazel, P., Todd, A.K., and Neidle, S. (2006). Quadruplex DNA: Sequence, topology and structure. Nucleic Acids Res. *34*, 5402–5415.

Busso, C.S., Iwakuma, T., and Izumi, T. (2009). Ubiquitination of mammalian AP endonuclease (APE1) regulated by the p53-MDM2 signaling pathway. Oncogene *28*, 1616–1625.

Calviello, A.K., Hirsekorn, A., Wurmus, R., Yusuf, D., and Ohler, U. (2019). Reproducible inference of transcription factor footprints in ATAC-seq and DNase- seq datasets using protocol-specific bias modeling. Genome Biol. *20*, 42.

Campbell, N.H., and Parkinson, G.N. (2007). Crystallographic studies of quadruplex nucleic acids. Methods *43*, 252–263.

Campbell, N., Collie, G.W., and Neidle, S. (2012). Crystallography of DNA and RNA G-Quadruplex nucleic acids and their ligand complexes. Curr. Protoc. Nucleic Acid Chem.

Chalupníková, K., Lattmann, S., Selak, N., Iwamoto, F., Fujiki, Y., and Nagamine, Y. (2008). Recruitment of the RNA helicase RHAU to stress granules via a unique RNA-binding domain. J. Biol. Chem. *283*, 35186–35198.

Chambers, V.S., Marsico, G., Boutell, J.M., Di Antonio, M., Smith, G.P., and Balasubramanian, S. (2015). High-throughput sequencing of DNA G-quadruplex structures in the human genome. Nat. Biotechnol. *33*, 877–881.

Chan, N., Ali, M., McCallum, G.P., Kumareswaran, R., Koritzinsky, M., Wouters, B.G., Wells, P.G., Gallinger, S., and Bristow, R.G. (2014). Hypoxia provokes base excision repair changes and a repair-deficient, mutator phenotype in colorectal cancer cells. Mol. Cancer Res. *12*, 1407–1415.

De Chaumont, F., Dallongeville, S., Chenouard, N., Hervé, N., Pop, S., Provoost, T., Meas-Yedid, V., Pankajakshan, P., Lecomte, T., Le Montagner, Y., et al. (2012). Icy: An open bioimage informatics platform for extended reproducible research. Nat. Methods *9*, 690–696.

Chen, F., Gao, X., Shilatifard, A., and Shilatifard, A. (2015a). Stably paused genes revealed through inhibition of transcription initiation by the TFIIH inhibitor triptolide. Genes Dev. *29*, 39–47.

Chen, F.X., Woodfin, A.R., Gardini, A., Rickels, R.A., Marshall, S.A., Smith, E.R., Shiekhattar, R., and Shilatifard, A. (2015b). PAF1, a Molecular Regulator of Promoter-Proximal Pausing by RNA Polymerase II. Cell *162*, 1003–1015.

Chen, F.X., Smith, E.R., and Shilatifard, A. (2018a). Born to run: Control of transcription elongation by RNA polymerase II. Nat. Rev. Mol. Cell Biol. *19*, 464–478.

Chen, H., Yan, Y., Davidson, T.L., Shinkai, Y., and Costa, M. (2006). Hypoxic stress induces dimethylated histone H3 lysine 9 through histone methyltransferase G9a in mammalian cells. Cancer Res. *66*, 9009–9016.

Chen, L., Chen, J.Y., Zhang, X., Gu, Y., Xiao, R., Shao, C., Tang, P., Qian, H., Luo, D., Li, H., et al. (2017). R-ChIP Using Inactive RNase H Reveals Dynamic Coupling of R-loops with Transcriptional Pausing at Gene Promoters. Mol. Cell *68*, 745-757.e5.

Chen, M.C., Murat, P., Abecassis, K., Ferre-D'Amare, A.R., and Balasubramanian, S. (2015c). Insights into the mechanism of a G-quadruplex-unwinding DEAH-box helicase. Nucleic Acids Res. *43*, 2223–2231.

Chen, M.C., Tippana, R., Demeshkina, N.A., Murat, P., Balasubramanian, S., Myong, S., and Ferré-D'amaré, A.R. (2018b). Structural basis of G-quadruplex unfolding by the DEAH/RHA helicase DHX36. Nature *558*, 465–483.

Chereji, R. V., Ramachandran, S., Bryson, T.D., and Henikoff, S. (2018). Precise genome-wide mapping of single nucleosomes and linkers in vivo. Genome Biol. *19*, 1–20.

Cheung, A.C.M., and Cramer, P. (2012). A movie of RNA polymerase II transcription. Cell *149*, 1431–1437.

Cheung, A.C.M., Sainsbury, S., and Cramer, P. (2011). Structural basis of initial RNA polymerase II transcription. EMBO J. *30*, 4755–4763.

Chimenti, F., Bizzarri, B., Maccioni, E., Secci, D., Bolasco, A., Chimenti, P., Fioravanti, R., Granese, A., Carradori, S., Tosi, F., et al. (2009). A novel histone acetyltransferase inhibitor modulating Gcn5 network: cyclopentylidene-[4-(4'-chlorophenyl)thiazol-2-yl)hydrazone. J. Med. Chem. *52*, 530–536.

Cho, W., Jayanth, N., English, B.P., Inoue, T., Andrews, J.O., Conway, W., Grimm, J.B., Spille, J., and Lavis, L.D. (2016). RNA Polymerase II cluster dynamics predict mRNA output in living cells. Elife *5*, 1–31.

Christiansen, J., Kofod, M., and Nielsen, F.C. (1994). A guanosine quadruplex and two stable hairpins flank a major cleavage site in insulin-like growth factor II mRNA. Nucleic Acids Res. *22*, 5709–5716.

De Cian, A., Guittat, L., Kaiser, M., Saccà, B., Amrane, S., Bourdoncle, A., Alberti, P., Teulade-Fichou, M.P., Lacroix, L., and Mergny, J.L. (2007a). Fluorescence-based melting assays for studying quadruplex ligands. Methods *42*, 183–195.

De Cian, A., DeLemos, E., Mergny, J.L., Teulade-Fichou, M.P., and Monchaud, D. (2007b). Highly efficient G-quadruplex recognition by bisquinolinium compounds. J. Am. Chem. Soc. *129*, 1856–1857.

Cogoi, S., and Xodo, L.E. (2006). G-quadruplex formation within the promoter of the KRAS proto-oncogene and its effect on transcription. Nucleic Acids Res. *34*, 2536–2549.

Cogoi, S., Paramasivam, M., Membrino, A., Yokoyama, K.K., and Xodo, L.E. (2010). The KRAS promoter responds to Myc-associated zinc finger and poly(ADP-ribose) polymerase 1 proteins, which recognize a critical quadruplex-forming GA-element. J. Biol. Chem. *285*, 22003–22016.

Conlon, E.G., Lu, L., Sharma, A., Yamazaki, T., Tang, T., Shneider, N.A., and Manley, J.L. (2016). The C9ORF72 GGGGCC expansion forms RNA G-quadruplex inclusions and sequesters hnRNP H to disrupt splicing in ALS brains. Elife *5*, 1–28.

Core, L.J., Waterfall, J.J., and Lis, J.T. (2008). Nascent RNA Sequencing Reveals Widespread Pausing and Divergent Initiation at Human Promoters. Science *322*, 1845–1848.

Cramer, P. (2019). Organization and regulation of gene transcription. Nature *573*, 45–54.

Cree, S.L., Fredericks, R., Miller, A., Pearce, F.G., Filichev, V., Fee, C., and Kennedy, M.A. (2016). DNA G-quadruplexes show strong interaction with DNA methyltransferases in vitro. FEBS Lett. *590*, 2870–2883.

Dahlin, J.L., Nelson, K.M., Strasser, J.M., Barsyte-Lovejoy, D., Szewczyk, M.M., Organ, S., Cuellar, M., Singh, G., Shrimp, J.H., Nguyen, N., et al. (2017). Assay interference and off-target liabilities of reported histone acetyltransferase inhibitors. Nat. Commun. *8*, 1527.

Dardenne, E., PolayEspinoza, M., Fattet, L., Germann, S., Lambert, M.P., Neil, H., Zonta, E., Mortada, H., Gratadou, L., Deygas, M., et al. (2014). RNA Helicases DDX5 and DDX17 Dynamically Orchestrate Transcription, miRNA, and Splicing Programs in Cell Differentiation. Cell Rep. *7*, 1900–1913.

Davie, J.R., and Moniwa, M. (2000). Control of chromatin remodeling. Crit. Rev. Eukaryot. Gene Expr. *10*, 303–325.

Dexheimer, T.S., Sun, D., and Hurley, L.H. (2006). Deconvoluting the structural and drug-recognition complexity of the G-quadruplex-forming region upstream of the bcl-2 P1 promoter. J. Am. Chem. Soc. *128*, 5404–5415.

Du, Z., Zhao, Y., and Li, N. (2008). Genome-wide analysis reveals regulatory role of G4 DNA in gene transcription. Genome Res. *18*, 516.

Dunham, I., Kundaje, A., Aldred, S.F., Collins, P.J., Davis, C.A., Doyle, F., Epstein, C.B., Frietze, S., Harrow, J., Kaul, R., et al. (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57–74.

Duquette, M.L., Handa, P., Vincent, J.A., Taylor, A.F., and Maizels, N. (2004). Intracellular transcription of G-rich DNAs induces formation of G-loops, novel structures containing G4

DNA. Genes Dev. *18*, 1618–1629.

Dutta, B., Yan, R., Lim, S.K., Tam, J.P., and Sze, S.K. (2014). Quantitative profiling of chromatome dynamics reveals a novel role for HP1BP3 in hypoxia-induced oncogenesis. Mol. Cell. Proteomics *13*, 3236–3249.

Eddy, J., and Maizels, N. (2006). Gene function correlates with potential for G4 DNA formation in the human genome. Nucleic Acids Res. *34*, 3887–3896.

Eddy, J., Vallur, A.C., Varma, S., Liu, H., Reinhold, W.C., Pommier, Y., and Maizels, N. (2011). G4 motifs correlate with promoter-proximal transcriptional pausing in human genes. Nucleic Acids Res. *39*, 4975–4983.

Egloff, S., Dienstbier, M., and Murphy, S. (2012). Updating the RNA polymerase CTD code: Adding gene-specific layers. Trends Genet. *28*, 333–341.

Eick, D., and Geyer, M. (2013). The RNA polymerase II carboxy-terminal domain (CTD) code. Chem. Rev. *113*, 8456–8490.

Erickson, B., Sheridan, R.M., Cortazar, M., and Bentley, D.L. (2018). Dynamic turnover of paused pol II complexes at human promoters. Genes Dev. *32*, 1215–1225.

Fazio, A., Briglia, M., Faggio, C., Alzoubi, K., and Lang, F. (2015). Stimulation of Suicidal Erythrocyte Death by Garcinol. Cell. Physiol. Biochem. *37*, 805–815.

Fernando, H., Rodriguez, R., and Balasubramanian, S. (2008). Selective recognition of a DNA G-quadruplex by an engineered antibody. Biochemistry *47*, 9365–9371.

Fleming, A.M., and Burrows, C.J. (2017). 8-Oxo-7,8-dihydro-2'-deoxyguanosine and abasic site tandem lesions are oxidation prone yielding hydantoin products that strongly destabilize duplex DNA. Org. Biomol. Chem. *15*, 8341–8353.

Fleming, A.M., Zhu, J., Ding, Y., and Burrows, C.J. (2017). 8-Oxo-7,8-dihydroguanine in the Context of a Gene Promoter G-Quadruplex Is an On-Off Switch for Transcription. ACS Chem. Biol. *12*, 2417–2426.

Fouquerel, E., Lormand, J., Bose, A., Lee, H.T., Kim, G.S., Li, J., Sobol, R.W., Freudenthal, B.D., Myong, S., and Opresko, P.L. (2016). Oxidative guanine base damage regulates human telomerase activity. Nat. Struct. Mol. Biol. *23*, 1092–1100.

Fouquerel, E., Barnes, R.P., Uttam, S., Watkins, S.C., Bruchez, M.P., and Opresko, P.L. (2019). Targeted and Persistent 8-Oxoguanine Base Damage at Telomeres Promotes Telomere Loss and Crisis. Mol. Cell *75*, 117-130.e6.

Franklin, R.E., and Gosling, R.G. (1953). Molecular Configuration in Sodium Thymonucleate. Nature *171*, 740–741.

Fujinaga, K., Irwin, D., Huang, Y., Taube, R., Kurosu, T., and Peterlin, B.M. (2004). P-TEFb

Phosphorylates RD and Dissociates Negative Effectors from the Transactivation Response Element. Mol. Cell. Biol. *24*, 787–795.

Gao, X., Lin, S.H., Ren, F., Li, J.T., Chen, J.J., Yao, C.B., Yang, H. Bin, Jiang, S.X., Yan, G.Q., Wang, D., et al. (2016). Acetate functions as an epigenetic metabolite to promote lipid synthesis under hypoxia. Nat. Commun. *7*, 11960.

Garant, J.M., Perreault, J.P., and Scott, M.S. (2017). Motif independent identification of potential RNA G-quadruplexes by G4RNA screener. Bioinformatics *33*, 3532–3537.

Gellert, I., Lipsett, M.N., and Davies, D.R. (1962). Helix formation by guanylic acid. Proc. Natl. Acad. Sci. U. S. A. *48*, 2013–2018.

Georgakopoulos-Soares, I., Morganella, S., Jain, N., Hemberg, M., and Nik-Zainal, S. (2017). Non-canonical secondary structures arising from non-B-DNA motifs are determinants of mutagenesis. Genome Res. 1–8.

Gerstein, M.B., Kundaje, A., Hariharan, M., Landt, S.G., Yan, K.K., Cheng, C., Mu, X.J., Khurana, E., Rozowsky, J., Alexander, R., et al. (2012). Architecture of the human regulatory network derived from ENCODE data. Nature *489*, 91–100.

Ghosh, A., Largy, E., and Gabelica, V. (2021). DNA G-quadruplexes for native mass spectrometry in potassium: a database of validated structures in electrospray-compatible conditions. Nucleic Acids Res. *49*, 2333–2345.

Ginno, P.A., Lott, P.L., Christensen, H.C., Korf, I., and Chédin, F. (2012). R-Loop Formation Is a Distinctive Characteristic of Unmethylated Human CpG Island Promoters. Mol. Cell *45*, 814–825.

Gkikopoulos, T., Schofield, P., Singh, V., Pinskaya, M., Mellor, J., Smolle, M., Workman, J.L., Barton, G.J., and Owen-Hughes, T. (2011). A role for Snf2-related nucleosome-spacing enzymes in genome-wide nucleosome organization. Science *333*, 1758–1760.

Gray, L.T., Vallur, A.C., Eddy, J., and Maizels, N. (2014). G quadruplexes are genomewide targets of transcriptional helicases XPB and XPD. Nat. Chem. Biol. *10*, 313–318.

Gregory, P.D., Wagner, K., and Hörz, W. (2001). Histone acetylation and chromatin remodeling. Exp. Cell Res. *265*, 195–202.

Guédin, A., Alberti, P., and Mergny, J.L. (2009). Stability of intramolecular quadruplexes: Sequence effects in the central loop. Nucleic Acids Res. *37*, 5559–5567.

Guédin, A., Gros, J., Alberti, P., and Mergny, J.L. (2010). How long is too long? Effects of loop size on G-quadruplex stability. Nucleic Acids Res. *38*, 7858–7868.

Guilbaud, G., Murat, P., Recolin, B., Campbell, B.C., Maiter, A., Sale, J.E., and Balasubramanian, S. (2017). Local epigenetic reprogramming induced by G-quadruplex

ligands. Nat. Chem.

Guo, J.U., and Bartel, D.P. (2016). RNA G-quadruplexes are globally unfolded in eukaryotic cells and depleted in bacteria. Science *353*, aaf5371–aaf5371.

Guo, K., Pourpak, A., Beetz-Rogers, K., Gokhale, V., Sun, D., and Hurley, L.H. (2007). Formation of pseudosymmetrical G-quadruplex and i-motif structures in the proximal promoter region of the RET oncogene. J. Am. Chem. Soc. *129*, 10220–10228.

Halder, R., Riou, J.F., Teulade-Fichou, M.P., Frickey, T., and Hartig, J.S. (2012). Bisquinolinium compounds induce quadruplex-specific transcriptome changes in HeLa S3 cell lines. BMC Res. Notes *5*, 138.

Han, H., Hurley, L.H., and Salazar, M. (1999). A DNA polymerase stop assay for G-quadruplex-interactive compounds. Nucleic Acids Res. *27*, 537–542.

Hänsel-Hertsch, R., Beraldi, D., Stefanie, L., Giovanni, M., Zyner, K., Parry, A., Di Antonio, M., Pike, J., Kimura, H., Narita, M., et al. (2016). G-quadruplex structures mark human regulatory chromatin. Nat Genet. *48*, 1267–1272.

Hänsel-Hertsch, R., Spiegel, J., Marsico, G., Tannahill, D., and Balasubramanian, S. (2018). Genome-wide mapping of endogenous G-quadruplex DNA structures by chromatin immunoprecipitation and high-throughput sequencing. Nat. Protoc. *13*, 551–564.

Hänsel-Hertsch, R., Simeone, A., Shea, A., Hui, W.W.I., Zyner, K.G., Marsico, G., Rueda, O.M., Bruna, A., Martin, A., Zhang, X., et al. (2020). Landscape of G-quadruplex DNA structural regions in breast cancer. Nat. Genet. *52*, 878–883.

Harada, Y., Ohara, O., Takatsuki, A., Itoh, H., Shimamoto, N., and Kinosita, K. (2001). Direct observation of DNA rotation during transcription by escherichia coli RNA polymerase. Nature *409*, 113–115.

Harlen, K.M., and Churchman, L.S. (2017). The code and beyond: Transcription regulation by the RNA polymerase II carboxy-terminal domain. Nat. Rev. Mol. Cell Biol. *18*, 263–273.

Hazel, P., Huppert, J., Balasubramanian, S., and Neidle, S. (2004). Loop-length-dependent folding of G-quadruplexes. J. Am. Chem. Soc. *126*, 16405–16415.

Heddi, B., Cheong, V.V., Martadinata, H., and Phan, A.T. (2015). Insights into G-quadruplex specific recognition by the DEAH-box helicase RHAU: Solution structure of a peptide-quadruplex complex. Proc. Natl. Acad. Sci. U. S. A. *112*, 9608–9613.

Hegde, M.L., Mantha, A.K., Hazra, T.K., Bhakat, K.K., Mitra, S., and Szczesny, B. (2012). Oxidative genome damage and its repair: Implications in aging and neurodegenerative diseases. Mech. Ageing Dev. *133*, 157–168.

Hegyi, H. (2015). Enhancer-promoter interaction facilitated by transiently forming G-

quadruplexes. Sci. Rep. *5*, 1–6.

Henderson, A., Wu, Y., Huang, Y.C., Chavez, E.A., Platt, J., Johnson, F.B., Brosh, R.M., Sen, D., and Lansdorp, P.M. (2014). Detection of G-quadruplex DNA in mammalian cells. Nucleic Acids Res. *42*, 860–869.

Herdy, B., Mayer, C., Varshney, D., Marsico, G., Murat, P., Taylor, C., D'Santos, C., Tannahill, D., and Balasubramanian, S. (2018). Analysis of NRAS RNA G-quadruplex binding proteins reveals DDX3X as a novel interactor of cellular G-quadruplex containing transcripts. Nucleic Acids Res. *46*, 11592–11604.

Hoffman, E.A., Frey, B.L., Smith, L.M., and Auble, D.T. (2015). Formaldehyde crosslinking: A tool for the study of chromatin complexes. J. Biol. Chem. *290*, 26404–26411.

Holder, I.T., and Hartig, J.S. (2014). A matter of location: Influence of G-quadruplexes on escherichia coli gene expression. Chem. Biol. *21*, 1511–1521.

Hong, S., and Kim, D. (2017). Computational characterization of chromatin domain boundary-associated genomic elements. Nucleic Acids Res. *45*, 10403–10414.

Hou, Y., Li, F., Zhang, R., Li, S., Liu, H., Qin, Z.S., and Sun, X. (2019). Integrative characterization of G-Quadruplexes in the three-dimensional chromatin structure. Epigenetics *14*, 894–911.

Huang, H., Zhang, J., Harvey, S.E., Hu, X., and Cheng, C. (2017). RNA G-quadruplex secondary structure promotes alternative splicing via the RNA-binding protein hnRNPF. Genes Dev. *31*, 2296–2309.

Huppert, J.L. (2008). Four-stranded nucleic acids: structure, function and targeting of G-quadruplexes. Chem. Soc. Rev. *37*, 1375.

Huppert, J.L., and Balasubramanian, S. (2005). Prevalence of quadruplexes in the human genome. Nucleic Acids Res. *33*, 2908–2916.

Huppert, J.L., and Balasubramanian, S. (2007). G-quadruplexes in promoters throughout the human genome. Nucleic Acids Res. *35*, 406–413.

Huppert, J.L., Bugaut, A., Kumari, S., and Balasubramanian, S. (2008). G-quadruplexes: The beginning and end of UTRs. Nucleic Acids Res. *36*, 6260–6268.

Hussain, T., Saha, D., Purohit, G., Kar, A., Kishore Mukherjee, A., Sharma, S., Sengupta, S., Dhapola, P., Maji, B., Vedagopuram, S., et al. (2017). Transcription regulation of CDKN1A (p21/CIP1/WAF1) by TRF2 is epigenetically controlled through the REST repressor complex. Sci. Rep. *7*, 1–13.

Hyun, K., Jeon, J., Park, K., and Kim, J. (2017). Writing, erasing and reading histone lysine methylations. Exp. Mol. Med. *49*, e324.

Iwase, S., Xiang, B., Ghosh, S., Ren, T., Lewis, P.W., Cochrane, J.C., Allis, C.D., Picketts, D.J., Patel, D.J., Li, H., et al. (2011). ATRX ADD domain links an atypical histone methylation recognition mechanism to human mental-retardation syndrome. Nat. Struct. Mol. Biol. *18*, 769–776.

Jeffers, V., Gao, H., Checkley, L.A., Liu, Y., Ferdig, M.T., and Sullivan, W.J. (2016). Garcinol Inhibits GCN5-Mediated Lysine Acetyltransferase Activity and Prevents Replication of the Parasite Toxoplasma gondii. Am. Soc. Microbiol. *60*, 2164–2170.

Kaczmarczyk, A., Meng, H., Ordu, O., Noort, J. van, and Dekker, N.H. (2020). Chromatin fibers stabilize nucleosomes under torsional stress. Nat. Commun. *11*, 1–12.

Kanoh, Y., Matsumoto, S., Fukatsu, R., Kakusho, N., Kono, N., Renard-Guillet, C., Masuda, K., Iida, K., Nagasawa, K., Shirahige, K., et al. (2015). Rif1 binds to G quadruplexes and suppresses replication over long distances. Nat. Struct. Mol. Biol. *22*, 889–897.

Kazemier, H.G., Paeschke, K., and Lansdorp, P.M. (2017). Guanine quadruplex monoclonal antibody 1H6 cross-reacts with restrained thymidine-rich single stranded DNA. Nucleic Acids Res. *45*, 5913–5919.

Ke, Y., Xu, Y., Chen, X., Feng, S., Liu, Z., Sun, Y., Yao, X., Li, F., Zhu, W., Gao, L., et al. (2017). 3D Chromatin Structures of Mature Gametes and Structural Reprogramming during Mammalian Embryogenesis. Cell *170*, 367-381.e20.

Kikin, O., D'Antonio, L., and Bagga, P.S. (2006). QGRS Mapper: A web-based server for predicting G-quadruplexes in nucleotide sequences. Nucleic Acids Res. *34*, 676–682.

Kim, J., Cheong, C., and Moore, P.B. (1991). Tetramerization of an RNA oligonucleotide containing a GGGG sequence. Nature *351*, 311–332.

Kim, M.Y., Vankayalapati, H., Shin-Ya, K., Wierzba, K., and Hurley, L.H. (2002). Telomestatin, a potent telomerase inhibitor that interacts quite specifically with the human telomeric intramolecular G-quadruplex. J. Am. Chem. Soc. *124*, 2098–2099.

Kim, N.W., Piatyszek, M.A., Prowse, K.R., Harley, C.B., West, M.D., Ho, P.L.C., Coviello, G.M., Wright, W.E., Weinrich, S.L., and Shay, J.W. (1994). Specific association of human telomerase activity with immortal cells and cancer. Science *266*, 2011–2015.

Kirmes, I., Szczurek, A., Prakash, K., Charapitsa, I., Heiser, C., Musheev, M., Schock, F., Fornalczyk, K., Ma, D., Birk, U., et al. (2015). A transient ischemic environment induces reversible compaction of chromatin. Genome Biol. *16*, 246.

Klemm, S.L., Shipony, Z., and Greenleaf, W.J. (2019). Chromatin accessibility and the regulatory epigenome. Nat. Rev. Genet. *20*, 207–220.

Kouzine, F., Liu, J., Sanford, S., Chung, H.J., and Levens, D. (2004). The dynamic response of

upstream DNA to transcription-generated torsional stress. Nat. Struct. Mol. Biol. *11*, 1092–1100.

Kouzine, F., Wojtowicz, D., Baranello, L., Yamane, A., Nelson, S., Resch, W., Kieffer-Kwon, K.-R., Benham, C.J., Casellas, R., Przytycka, T.M., et al. (2017). Permanganate/S1 Nuclease Footprinting Reveals Non-B DNA Structures with Regulatory Potential across a Mammalian Genome. Cell Syst. 1–13.

Krebs, A.R., Imanci, D., Hoerner, L., Gaidatzis, D., Burger, L., and Schübeler, D. (2017). Genome-wide Single-Molecule Footprinting Reveals High RNA Polymerase II Turnover at Paused Promoters. Mol. Cell *67*, 411-422.e4.

Kruisselbrink, E., Guryev, V., Brouwer, K., Pontier, D.B., Cuppen, E., and Tijsterman, M. (2008). Mutagenic Capacity of Endogenous G4 DNA Underlies Genome Instability in FANCJ-Defective C. elegans. Curr. Biol. *18*, 900–905.

Kudlicki, A.S. (2016). G-quadruplexes involving both strands of genomic DNA are highly abundant and colocalize with functional sites in the human genome. PLoS One *11*, 1–12.

Kuehner, J.N., Pearson, E.L., and Moore, C. (2011). Unravelling the means to an end: RNA polymerase II transcription termination. Nat. Rev. Mol. Cell Biol. *12*, 283–294.

Kumar, P., Yadav, V.K., Baral, A., Kumar, P., Saha, D., and Chowdhury, S. (2011). Zinc-finger transcription factors are associated with guanine quadruplex motifs in human, chimpanzee, mouse and rat promoters genome-wide. Nucleic Acids Res. *39*, 8005–8016.

Kumari, S., Bugaut, A., Huppert, J.L., and Balasubramanian, S. (2007). An RNA G-quadruplex in the 5′ UTR of the NRAS proto-oncogene modulates translation. Nat. Chem. Biol. *3*, 218–221.

Kwok, C.K., and Balasubramanian, S. (2015). Targeted detection of G-quadruplexes in cellular RNAs. Angew. Chem. Int. Ed. *54*, 6751–6754.

Kwok, C.K., Marsico, G., Sahakyan, A.B., Chambers, V.S., and Balasubramanian, S. (2016a). rG4-seq reveals widespread formation of G-quadruplex structures in the human transcriptome. Nat. Methods *13*, 841–844.

Kwok, C.K., Sahakyan, A.B., and Balasubramanian, S. (2016b). Structural Analysis using SHALiPE to Reveal RNA G-Quadruplex Formation in Human Precursor MicroRNA. Angew. Chemie - Int. Ed. *55*, 8958–8961.

Laitem, C., Zaborowska, J., Isa, N.F., Kufs, J., Dienstbier, M., and Murphy, S. (2015). CDK9 inhibitors define elongation checkpoints at both ends of RNA polymerase II-transcribed genes. Nat. Struct. Mol. Biol. *22*, 396–403.

Lam, E.Y.N., Beraldi, D., Tannahill, D., and Balasubramanian, S. (2013). G-quadruplex

structures are stable and detectable in human genomic DNA. Nat. Commun. *4*, 1796–1798.
Längst, G., and Manelyte, L. (2015). Chromatin remodelers: From function to dysfunction. Genes (Basel). *6*, 299–324.
Law, M.J., Lower, K.M., Voon, H.P.J., Hughes, J.R., Garrick, D., Viprakasit, V., Mitson, M., De Gobbi, M., Marra, M., Morris, A., et al. (2010). ATR-X syndrome protein targets tandem repeats and influences allele-specific expression in a size-dependent manner. Cell *143*, 367–378.
Lefebvre, J., Guetta, C., Poyer, F., Mahuteau-Betzer, F., and Teulade-Fichou, M.P. (2017). Copper–Alkyne Complexation Responsible for the Nucleolar Localization of Quadruplex Nucleic Acid Drugs Labeled by Click Reactions. Angew. Chemie - Int. Ed. *56*, 11365–11369.
Lemmens, B., Van Schendel, R., and Tijsterman, M. (2015). Mutagenic consequences of a single G-quadruplex demonstrate mitotic inheritance of DNA replication fork barriers. Nat. Commun. *6*, 1–8.
Li, G., and Reinberg, D. (2011). Chromatin higher-order structures and gene regulation. Curr. Opin. Genet. Dev. *21*, 175–186.
Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.
Li, Q., Xiang, J.F., Yang, Q.F., Sun, H.X., Guan, A.J., and Tang, Y.L. (2013). G4LDB: A database for discovering and studying G-quadruplex ligands. Nucleic Acids Res. *41*, 1115–1123.
Liao, C.H., Sang, S., Ho, C.T., and Lin, J.K. (2005). Garcinol modulates tyrosine phosphorylation of FAK and subsequently induces apoptosis through down-regulation of Src, ERK, and Akt survival signaling in human colon cancer cells. J. Cell. Biochem. *96*, 155–169.
Lim, K.W., Amrane, S., Bouaziz, S., Xu, W., Mu, Y., Patel, D.J., Luu, K.N., and Phan, A.T. (2009). Structure of the human telomere in K + solution: A stable basket-type G-quadruplex with only two G-tetrad layers. J. Am. Chem. Soc. *131*, 4301–4309.
Liu, L., and Gerson, S.L. (2004). Therapeutic impact of methoxyamine: blocking repair of abasic sites in the base excision repair pathway. Curr. Opin. Investig. Drugs *5*, 623–627.
Liu, L.F., and Wang, J.C. (1987). Supercoiling of the DNA template during transcription. Proc. Natl. Acad. Sci. U. S. A. *84*, 7024–7027.
Liu, H.Y., Zhao, Q., Zhang, T.P., Wu, Y., Xiong, Y.X., Wang, S.K., Ge, Y.L., He, J.H., Lv, P., Ou, T.M., et al. (2016). Conformation Selective Antibody Enables Genome Profiling and Leads to Discovery of Parallel G-Quadruplex in Human Telomeres. Cell Chem. Biol. *23*, 1261–1270.
Liu, X., Zhang, Y., Chen, Y., Li, M., Zhou, F., Li, K., Cao, H., Ni, M., Liu, Y., Gu, Z., et al. (2017).

In Situ Capture of Chromatin Interactions by Biotinylated dCas9. Cell *170*, 1028-1043.e19.

Lorenzini, P.A., Chew, R.S.E., Tan, C.W., Yong, J.Y., Zhang, F., Zheng, J., and Roca, X. (2019). Human PRPF40B regulates hundreds of alternative splicing targets and represses a hypoxia expression signature. RNA *25*, 905–920.

Lyons, S.M., Achorn, C., Kedersha, N.L., Anderson, P.J., and Ivanov, P. (2016). YB-1 regulates tiRNA-induced Stress Granule formation but not translational repression. Nucleic Acids Res. *44*, 6949–6960.

Lyons, S.M., Gudanis, D., Coyne, S.M., Gdaniec, Z., and Ivanov, P. (2017). Identification of functional tetramolecular RNA G-quadruplexes derived from transfer RNAs. Nat. Commun. *8*, 1127.

Madireddy, A., Purushothaman, P., Loosbroock, C.P., Robertson, E.S., Schildkraut, C.L., and Verma, S.C. (2016). G-quadruplex-interacting compounds alter latent DNA replication and episomal persistence of KSHV. Nucleic Acids Res. *44*, 3675–3694.

De Magis, A., Manzo, S.G., Russo, M., Marinello, J., Morigi, R., Sordet, O., and Capranico, G. (2019). DNA damage and genome instability by G-quadruplex ligands are mediated by R loops in human cancer cells. Proc. Natl. Acad. Sci. U. S. A. *116*, 816–825.

Makowski, M.M., Gräwe, C., Foster, B.M., Nguyen, N. V., Bartke, T., and Vermeulen, M. (2018). Global profiling of protein-DNA and protein-nucleosome binding affinities using quantitative mass spectrometry. Nat. Commun. *9*, 1653.

Mao, S.Q., Ghanbarian, A.T., Spiegel, J., Martínez Cuesta, S., Beraldi, D., Di Antonio, M., Marsico, G., Hänsel-Hertsch, R., Tannahill, D., and Balasubramanian, S. (2018). DNA G-quadruplex structures mold the DNA methylome. Nat. Struct. Mol. Biol. *25*, 951–957.

Marchetti, C., Zyner, K.G., Ohnmacht, S.A., Robson, M., Haider, S.M., Morton, J.P., Marsico, G., Vo, T., Laughlin-Toth, S., Ahmed, A.A., et al. (2018). Targeting Multiple Effector Pathways in Pancreatic Ductal Adenocarcinoma with a G-Quadruplex-Binding Small Molecule. J. Med. Chem. *61*, 2500–2517.

Marcu, M.G., Jung, Y.-J., Lee, S., Chung, E.-J., Lee, M.-J., Trepel, J., and Neckers, L. (2006). Curcumin is an inhibitor of p300 histone acetylatransferase. Med. Chem. *2*, 169–174.

Marshall, N.F., and Price, D.H. (1995). Purification of P-TEFb, a transcription factor required for the transition into productive elongation. J. Biol. Chem. *270*, 12335–12338.

Marshall, N.F., Peng, J., Xie, Z., and Price, D.H. (1996). Control of RNA polymerase II elongation potential by a novel carboxyl- terminal domain kinase. J. Biol. Chem. *271*, 27176–27183.

Marsico, G., Chambers, V.S., Sahakyan, A.B., McCauley, P., Boutell, J.M., Antonio, M. Di, and

Balasubramanian, S. (2019). Whole genome experimental maps of DNA G-quadruplexes in multiple species. Nucleic Acids Res. *47*, 3862–3874.

Martadinata, H., and Phan, A.T. (2009). Structure of propeller-type parallel-stranded RNA G-quadruplexes, formed by human telomeric RNA sequences in K+ solution. J. Am. Chem. Soc. *131*, 2570–2579.

Martin, M. (2011). Cutadapt Removes Adapter Sequences From High-Throughput Sequencing Reads. EMBnet.Journal *17*, 10–12.

Mathad, R.I., Hatzakis, E., Dai, J., and Yang, D. (2011). C-MYC promoter G-quadruplex formed at the 5'-end of NHE III 1 element: Insights into biological relevance and parallel-stranded G-quadruplex stability. Nucleic Acids Res. *39*, 9023–9033.

McLuckie, K.I.E., Di Antonio, M., Zecchini, H., Xian, J., Caldas, C., Krippendorff, B.F., Tannahill, D., Lowe, C., and Balasubramanian, S. (2013). G-quadruplex DNA as a molecular target for induced synthetic lethality in cancer cells. J. Am. Chem. Soc. *135*, 9640–9643.

Mergny, J.L., and Lacroix, L. (2009). UV melting of G-quadruplexes. Curr. Protoc. Nucleic Acid Chem. 1–15.

Mergny, J.L., Li, J., Lacroix, L., Amrane, S., and Chaires, J.B. (2005). Thermal difference spectra: A specific signature for nucleic acid structures. Nucleic Acids Res. *33*, 1–6.

Mergny, J.Y., Phan, A., and Lacroix, L. (1998). Following G-quartet formation by UV-spectroscopy. Fed. Eur. Biochem. Soc. *435*, 74–78.

Mirihana Arachchilage, G., Hetti Arachchilage, M., Venkataraman, A., Piontkivska, H., and Basu, S. (2019). Stable G-quadruplex enabling sequences are selected against by the context-dependent codon bias. Gene *696*, 149–161.

Moye, A.L., Porter, K.C., Cohen, S.B., Phan, T., Zyner, K.G., Sasaki, N., Lovrecz, G.O., Beck, J.L., and Bryan, T.M. (2015). Telomeric G-quadruplexes are a substrate and site of localization for human telomerase. Nat. Commun. *6*.

Mukundan, V.T., and Phan, A.T. (2013). Bulges in G-quadruplexes: Broadening the definition of G-quadruplex-forming sequences. J. Am. Chem. Soc. *135*, 5017–5028.

Müller, S., Kumari, S., Rodriguez, R., and Balasubramanian, S. (2010). Small-molecule-mediated G-quadruplex isolation from human cells. Nat. Chem. *2*, 1095–1098.

Murat, P., and Balasubramanian, S. (2014). Existence and consequences of G-quadruplex structures in DNA. Curr. Opin. Genet. Dev. *25*, 22–29.

Murat, P., Marsico, G., Herdy, B., Ghanbarian, A., Portella, G., and Balasubramanian, S. (2018). RNA G-quadruplexes at upstream open reading frames cause DHX36- and DHX9-dependent translation of human mRNAs. Genome Biol. *19*, 1–24.

Muz, B., de la Puente, P., Azab, F., and Azab, A.K. (2015). The role of hypoxia in cancer progression, angiogenesis, metastasis, and resistance to therapy. Hypoxia *3*, 83–92.

Nag, A., Narsinh, K., and Martinson, H.G. (2007). The poly(A)-dependent transcriptional pause is mediated by CPSF acting on the body of the polymerase. Nat. Struct. Mol. Biol. *14*, 662–669.

Neidle, S. (2010). Human telomeric G-quadruplex: The current status of telomeric G-quadruplexes as therapeutic targets in human cancer. FEBS J. *277*, 1118–1125.

Ni, Z., Saunders, A., Fuda, N.J., Yao, J., Suarez, J.-R., Webb, W.W., and Lis, J.T. (2008). P-TEFb Is Critical for the Maturation of RNA Polymerase II into Productive Elongation In Vivo. Mol. Cell. Biol. *28*, 1161–1170.

Ninova, M., Tóth, K.F., and Aravin, A.A. (2019). The control of gene expression and cell identity by H3K9 trimethylation. Dev. *146*, dev181180.

Niu, K., Xiang, L., Jin, Y., Peng, Y., Wu, F., Tang, W., Zhang, X., Deng, H., Xiang, H., Li, S., et al. (2019). Identification of LARK as a novel and conserved G-quadruplex binding protein in invertebrates and vertebrates. Nucleic Acids Res. *47*, 7306–7320.

Paeschke, K., Simonsson, T., Postberg, J., Rhodes, D., and Lipps, H.J. (2005). Telomere end-binding proteins control the formation of G-quadruplex DNA structures in vivo. Nat. Struct. Mol. Biol. *12*, 847–854.

Paeschke, K., Capra, J.A., and Zakian, V.A. (2011). DNA Replication through G-Quadruplex Motifs Is Promoted by the Saccharomyces cerevisiae Pif1 DNA Helicase. Cell *145*, 678–691.

Pan, M.H., Chang, W.L., Lin-Shiau, S.Y., Ho, C.T., and Lin, J.K. (2001). Induction of apoptosis by garcinol and curcumin through cytochrome c release and activation of caspases in human leukemia HL-60 cells. J. Agric. Food Chem. *49*, 1464–1474.

Papadopoulou, C., Guilbaud, G., Schiavone, D., and Sale, J.E. (2015). Nucleotide Pool Depletion Induces G-Quadruplex-Dependent Perturbation of Gene Expression. Cell Rep. *13*, 2491–2503.

Park, M., Upton, D., Blackmon, M., Dixon, V., Craver, S., Neal, D., and Perkins, D. (2018). Anacardic acid inhibits pancreatic cancer cell growth, and potentiates chemotherapeutic effect by Chmp1A - ATM - p53 signaling pathway. BMC Complement. Altern. Med. *18*, 1–10.

Parkinson, G.N., Lee, M.P.H., and Neidle, S. (2002). Crystal structure of parallel quadruplexes from human telomeric DNA. Nature *417*, 876–880.

Parthun, M.R. (2007). Hat1: The emerging cellular roles of a type B histone acetyltransferase. Oncogene *26*, 5319–5328.

Phan, A.T. (2010). Human telomeric G-quadruplex: Structures of DNA and RNA sequences.

FEBS J. *277*, 1107–1117.

Phan, A.T., Kuryavyi, V., Burge, S., Neidle, S., and Patel, D.J. (2007). Structure of an unprecedented G-quadruplex scaffold in the human c-kit promoter. J. Am. Chem. Soc. *129*, 4386–4392.

Piazza, A., Boulé, J.B., Lopes, J., Mingo, K., Largy, E., Teulade-Fichou, M.P., and Nicolas, A. (2010). Genetic instability triggered by G-quadruplex interacting Phen-DC compounds in Saccharomyces cerevisiae. Nucleic Acids Res. *38*, 4337–4348.

Piazza, A., Adrian, M., Samazan, F., Heddi, B., Hamon, F., Serero, A., Lopes, J., Teulade-Fichou, M.-P., Phan, A.T., and Nicolas, A. (2015). Short loop length and high thermal stability determine genomic instability induced by G-quadruplex-forming minisatellites. EMBO J. *34*, e201490702.

Piazza, A., Cui, X., Adrian, M., Samazan, F., Heddi, B., Phan, A.T., and Nicolas, A.G. (2017). Non-canonical G-quadruplexes cause the hCEB1 minisatellite instability in Saccharomyces cerevisiae. Elife *6*, 1–21.

Pines, A., Perrone, L., Bivi, N., Romanello, M., Damante, G., Gulisano, M., Kelley, M.R., Quadrifoglio, F., and Tell, G. (2005). Activation of APE1/Ref-1 is dependent on reactive oxygen species generated after purinergic receptor stimulation by ATP. Nucleic Acids Res. *33*, 4379–4394.

Poirier, M.G., Bussiek, M., Langowski, J., and Widom, J. (2008). Spontaneous Access to DNA Target Sites in Folded Chromatin Fibers. J. Mol. Biol. *379*, 772–786.

Price, D.H. (2000). P-TEFb, a Cyclin-Dependent Kinase Controlling Elongation by RNA Polymerase II. Mol. Cell. Biol. *20*, 2629–2634.

Qin, Y., Rezler, E.M., Gokhale, V., Sun, D., and Hurley, L.H. (2007). Characterization of the G-quadruplexes in the duplex nuclease hypersensitive element of the PDGF-A promoter and modulation of PDGF-A promoter activity by TMPyP4. Nucleic Acids Res. *35*, 7698–7713.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

Quinodoz, M., Gobet, C., Naef, F., and Gustafson, K.B. (2014). Characteristic bimodal profiles of RNA polymerase II at thousands of active mammalian promoters. Genome Biol. *15*, R85.

Raiber, E.A., Kranaster, R., Lam, E., Nikan, M., and Balasubramanian, S. (2012). A non-canonical DNA structure is a binding motif for the transcription factor SP1 in vitro. Nucleic Acids Res. *40*, 1499–1508.

Ramírez, F., Dündar, F., Diehl, S., Grüning, B.A., and Manke, T. (2014). DeepTools: A flexible platform for exploring deep-sequencing data. Nucleic Acids Res. *42*, W187-W191.

Rao, S.S.P., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., et al. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell *159*, 1665–1680.

Reeg, S., Castro, J.P., Hugo, M., and Grune, T. (2020). Accumulation of polyubiquitinated proteins: A consequence of early inactivation of the 26S proteasome. Free Radic. Biol. Med. *160*, 293–302.

Revyakin, A., Ebright, R.H., and Strick, T.R. (2004). Promoter unwinding and promoter clearance by RNA polymerase: Detection by single-molecule DNA nanomanipulation. Proc. Natl. Acad. Sci. U. S. A. *101*, 4776–4780.

Richmond, T.J., Finch, J.T., and Klug, A. (1986). The structure of the nucleosome core particle. Biochem. Soc. Trans. *14*, 221–221.

Robichon, C., Luo, J., Causey, T.B., Benner, J.S., and Samuelson, J.C. (2011). Engineering Escherichia coli BL21(DE3) derivative strains to minimize E. coli Protein contamination after purification by immobilized metal affinity chromatography. Appl. Environ. Microbiol. *77*, 4634–4646.

Robinson, P.J.J., An, W., Routh, A., Martino, F., Chapman, L., Roeder, R.G., and Rhodes, D. (2008). 30 nm Chromatin Fibre Decompaction Requires both H4-K16 Acetylation and Linker Histone Eviction. J. Mol. Biol. *381*, 816–825.

Rocha, S. (2007). Gene regulation under low oxygen: holding your breath for transcription. Trends Biochem. Sci. *32*, 389–397.

Rodriguez, R., Miller, K.M., Forment, J. V, Bradshaw, C.R., Nikan, M., Britton, S., Oelschlaegel, T., Xhemalce, B., Balasubramanian, S., and Jackson, S.P. (2012). Small-molecule–induced DNA damage identifies alternative DNA structures in human genes. Nat. Chem. Biol. *8*, 301–310.

Rosonina, E., Kaneko, S., and Manley, J.L. (2006). Terminating the transcript: Breaking up is hard to do. Genes Dev. *20*, 1050–1056.

Roychoudhury, S., Pramanik, S., Harris, H.L., Tarpley, M., Sarkar, A., Spagnol, G., Sorgen, P.L., Chowdhury, D., Band, V., Klinkebiel, D., et al. (2020). Endogenous oxidized DNA bases and APE1 regulate the formation of G-quadruplex structures in the genome. Proc. Natl. Acad. Sci. *117*, 11409–11420.

Saffi, J., Agnoletto, M.H., Guecheva, T.N., Batista, L.F.Z., Carvalho, H., Henriques, J.A.P., Stary, A., Menck, C.F.M., and Sarasin, A. (2010). Effect of the anti-neoplastic drug doxorubicin on XPD-mutated DNA repair-deficient human cells. DNA Repair (Amst). *9*, 40–47.

Saha, D., Singh, A., Hussain, T., Srivastava, V., Sengupta, S., Kar, A., Dhapola, P., Dhople, V.,

Ummanni, R., and Chowdhury, S. (2017). Epigenetic suppression of human telomerase (hTERT) is mediated by the metastasis suppressor NME2 in a G-quadruplex– dependent fashion. J. Biol. Chem. *292*, 15205–15215.

Saksouk, N., Simboeck, E., and Déjardin, J. (2015). Constitutive heterochromatin formation and transcription in mammals. Epigenetics and Chromatin *8*, 1–17.

Salvati, E., Scarsella, M., Porru, M., Rizzo, A., Iachettini, S., Tentori, L., Graziani, G., D'Incalci, M., Stevens, M.F.G., Orlandi, A., et al. (2010). PARP1 is activated at telomeres upon G4 stabilization: Possible target for telomere-based therapy. Oncogene *29*, 6280–6293.

Santos-Pereira, J.M., and Aguilera, A. (2015). R loops: New modulators of genome dynamics and function. Nat. Rev. Genet. *16*, 583–597.

Sarkies, P., Reams, C., Simpson, L.J., and Sale, J.E. (2010). Epigenetic Instability due to Defective Replication of Structured DNA. Mol. Cell *40*, 703–713.

Sauer, M., Juranek, S.A., Marks, J., De Magis, A., Kazemier, H.G., Hilbig, D., Benhalevy, D., Wang, X., Hafner, M., and Paeschke, K. (2019). DHX36 prevents the accumulation of translationally inactive mRNAs with G4-structures in untranslated regions. Nat. Commun. *10*, 2421.

Schaffitzel, C., Berger, I., Postberg, J., Hanes, J., Lipps, H.J., and Plückthun, A. (2001). In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with Stylonychia lemnae macronuclei. Proc. Natl. Acad. Sci. U. S. A. *98*, 8572–8577.

Schoeftner, S., and Blasco, M.A. (2008). Developmentally regulated transcription of mammalian telomeres by DNA-dependent RNA polymerase II. Nat. Cell Biol. *10*, 228–236.

Semenza, G.L. (2010). Defining the role of hypoxia-inducible factor 1 in cancer biology and therapeutics. Oncogene *29*, 625–634.

Sen, D., and Gilbert, W. (1998). Formation of parallel four-stranded complexes by guanine-rich motif for meiosis. Nature *334*, 364–366.

Sengupta, P., Bhattacharya, A., Sa, G., Das, T., and Chatterjee, S. (2019). Truncated G-Quadruplex Isomers Cross-Talk with the Transcription Factors to Maintain Homeostatic Equilibria in c-MYC Transcription. Biochemistry *58*, 1975–1991.

Shahid, R., Bugaut, A., and Balasubramanian, S. (2010). The BCL-2 5′ untranslated region contains an RNA G-quadruplex-forming motif that modulates protein expression. Biochemistry *49*, 8300–8306.

Shay, J.W., and Wright, W.E. (2019). Telomeres and telomerase: three decades of progress. Nat. Rev. Genet. *20*, 299–309.

Shin-ya, K., Wierzba, K., Matsuo, K., Ohtani, T., Yamada, Y., Furihata, K., Hayakawa, Y., and

Seto, H. (2001). Telomestatin, a novel telomerase inhibitor from Streptomyces anulatus [17]. J. Am. Chem. Soc. *123*, 1262–1263.

Shivalingam, A., Izquierdo, M.A., Marois, A. Le, Vyšniauskas, A., Suhling, K., Kuimova, M.K., and Vilar, R. (2015). The interactions between a small molecule and G-quadruplexes are visualized by fluorescence lifetime imaging microscopy. Nat. Commun. *6*, 8178.

Siddiqui-Jain, A., Grand, C.L., Bearss, D.J., and Hurley, L.H. (2002). Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. Proc. Natl. Acad. Sci. U. S. A. *99*, 11593–11598.

Simonsson, T., Pecinka, P., and Kubista, M. (1998). DNA tetraplex formation in the control region of c-myc. Nucleic Acids Res. *26*, 1167–1172.

Smith, J.S., Chen, Q., Yatsunyk, L.A., Nicoludis, J.M., Garcia, M.S., Kranaster, R., Balasubramanian, S., Monchaud, D., Teulade-Fichou, M.P., Abramowitz, L., et al. (2011). Rudimentary G-quadruplex-based telomere capping in Saccharomyces cerevisiae. Nat. Struct. Mol. Biol. *18*, 478–486.

Sollier, J., and Cimprich, K.A. (2015). Breaking bad: R-loops and genome integrity. Trends Cell Biol. *25*, 514–522.

Spiegel, J., Cuesta, S.M., Adhikari, S., Hänsel-Hertsch, R., Tannahill, D. and Balasubramanian, S. (2021). G-quadruplexes are transcription factor binding hubs in human chromatin. Genome Biol 22, 117.

Stegle, O., Payet, L., Mergny, J.L., MacKay, D.J.C., and Huppert, J.L. (2009). Predicting and understanding the stability of G-quadruplexes. Bioinformatics *25*, 374–382.

Summers, P.A., Lewis, B.W., Gonzalez-Garcia, J., Porreca, R.M., Lim, A.H.M., Cadinu, P., Martin-Pintado, N., Mann, D.J., Edel, J.B., Vannier, J.B., et al. (2021). Visualising G-quadruplex DNA dynamics in live cells by fluorescence lifetime imaging microscopy. Nat. Commun. *12*, 1–11.

Sun, D., and Hurley, L.H. (2010). Biochemical Techniques for the Characterization of G-Quadruplex Structures: EMSA, DMS Footprinting, and DNA Polymerase Stop Assay. *608*, 65–79.

Sun, D., Thompson, B., Cathers, B.E., Salazar, M., Kerwin, S.M., Trent, J.O., Jenkins, T.C., Neidle, S., and Hurley, L.H. (1997). Inhibition of human telomerase by a G-Quadruplex-Interactive compound. J. Med. Chem. *40*, 2113–2116.

Sun, D., Guo, K., and Shin, Y.-J. (2011). Evidence of the formation of G-quadruplex structures in the promoter region of the human vascular endothelial growth factor gene. Nucleic Acids Res. *39*, 1256–1265.

Sun, H., Yabuki, A., and Maizels, N. (2001). A human nuclease specific for G4 DNA. Proc. Natl. Acad. Sci. U. S. A. *98*, 12444–12449.

Svaren, J., Klebanow, E., Sealy, L., and Chalkley, R. (1994). Analysis of the competition between nucleosome formation and transcription factor binding. J. Biol. Chem. *269*, 9335–9344.

Swygert, S.G., and Peterson, C.L. (2014). Chromatin dynamics: Interplay between remodeling enzymes and histone modifications. Biochim. Biophys. Acta - Gene Regul. Mech. *1839*, 728–736.

Takahama, K., Takada, A., Tada, S., Shimizu, M., Sayama, K., Kurokawa, R., and Oyoshi, T. (2013). Regulation of telomere length by G-quadruplex telomere DNA- and TERRA-binding protein TLS/FUS. Chem. Biol. *20*, 341–350.

Tao, E.-W., Cheng, W.Y., Li, W.-L., Yu, J., and Gao, Q.-Y. (2020). tiRNAs: A novel class of small noncoding RNAs that helps cells respond to stressors and plays roles in cancer progression. J. Cell. Physiol. *235*, 683–690.

Tell, G., Damante, G., Caldwell, D., and Kelley, M.R. (2005). The Intracellular Localization of APE 1 /Ref-1: More Than a Passive Phenomenon? Antioxid. Redox Signal. *7*, 367–378.

Titov, D. V., Gilman, B., He, Q.L., Bhat, S., Low, W.K., Dang, Y., Smeaton, M., Demain, A.L., Miller, P.S., Kugel, J.F., et al. (2011). XPB, a subunit of TFIIH, is a target of the natural product triptolide. Nat. Chem. Biol. *7*, 182–188.

Tjeertes, J. V., Miller, K.M., and Jackson, S.P. (2009). Screen for DNA-damage-responsive histone modifications identifies H3K9Ac and H3K56Ac in human cells. EMBO J. *28*, 1878–1889.

Todd, A.K., and Neidle, S. (2008). The relationship of potential G-quadruplex sequences in cis-upstream regions of the human genome to SP1-binding elements. Nucleic Acids Res. *36*, 2700–2704.

Todd, A.K., Johnston, M., and Neidle, S. (2005). Highly prevalent putative quadruplex sequence motifs in human DNA. Nucleic Acids Res. *33*, 2901–2907.

Trisciuoglio, D., Ragazzoni, Y., Pelosi, A., Desideri, M., Carradori, S., Gabellini, C., Maresca, G., Nescatelli, R., Secci, D., Bolasco, A., et al. (2012). CPTH6, a thiazole derivative, induces histone hypoacetylation and apoptosis in human leukemia cells. Clin. Cancer Res. *18*, 475–486.

Valentin-Vega, Y.A., Wang, Y.D., Parker, M., Patmore, D.M., Kanagaraj, A., Moore, J., Rusch, M., Finkelstein, D., Ellison, D.W., Gilbertson, R.J., et al. (2016). Cancer-associated DDX3X mutations drive stress granule assembly and impair global translation. Sci. Rep. *6*, 1–16.

Varizhuk, A., Ischenko, D., Tsvetkov, V., Novikov, R., Kulemin, N., Kaluzhny, D., Vlasenok, M., Naumov, V., Smirnov, I., and Pozmogova, G. (2017). The expanding repertoire of G4 DNA structures. Biochimie *135*, 54–62.

Varshney, D., Vavrova-Anderson, J., Oler, A.J., Cowling, V.H., Cairns, B.R., and White, R.J. (2015). SINE transcription by RNA polymerase III is suppressed by histone methylation but not by DNA methylation. Nat. Commun. *6*, 6569.

del Villar-Guerra, R., Trent, J.O., and Chaires, J.B. (2018). G-Quadruplex Secondary Structure Obtained from Circular Dichroism Spectroscopy. Angew. Chemie - Int. Ed. *57*, 7171–7175.

Voigt, P., Tee, W.W., and Reinberg, D. (2013). A double take on bivalent promoters. Genes Dev. *27*, 1318–1338.

Vos, S.M., Farnung, L., Boehning, M., Wigge, C., Linden, A., Urlaub, H., and Cramer, P. (2018). Structure of activated transcription complex Pol II–DSIF–PAF–SPT6. Nature *560*, 607–612.

Wada, T., Takagi, T., Yamaguchi, Y., Ferdous, A., Imai, T., Hirose, S., Sugimoto, S., Yano, K., Hartzog, G.A., Winston, F., et al. (1998). DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs. Genes Dev. *12*, 343–356.

Wang, Y., Lu, J. jian, He, L., and Yu, Q. (2011). Triptolide (TPL) inhibits global transcription by inducing proteasome-dependent degradation of RNA polymerase II (Pol II). PLoS One *6*.

Wang, Y., Yang, J., Wild, A.T., Wu, W.H., Shah, R., Danussi, C., Riggins, G.J., Kannan, K., Sulman, E.P., Chan, T.A., et al. (2019). G-quadruplex DNA drives genomic instability and represents a targetable molecular abnormality in ATRX-deficient malignant glioma. Nat. Commun. *10*, 943.

Wang, Z., Zang, C., Cui, K., Schones, D.E., Barski, A., Peng, W., and Zhao, K. (2009). Genome-wide Mapping of HATs and HDACs Reveals Distinct Functions in Active and Inactive Genes. Cell *138*, 1019–1031.

Wanrooij, P.H., Uhler, J.P., Shi, Y., Westerlund, F., Falkenberg, M., and Gustafsson, C.M. (2012). A hybrid G-quadruplex structure formed between RNA and DNA explains the extraordinary stability of the mitochondrial R-loop. Nucleic Acids Res. *40*, 10334–10344.

Watson, J.D., and Crick, F.H.C. (1953). Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. Nature *171*, 740–741.

Webba da Silva, M. (2007). NMR methods for studying quadruplex nucleic acids. Methods *43*, 264–277.

Wei, D., Husby, J., and Neidle, S. (2015). Flexibility and structural conservation in a c-KIT G-quadruplex. Nucleic Acids Res. *43*, 629–644.

Weitzmann, M.N., Woodford, K.J., and Usdin, K. (1996). The development and use of a DNA polymerase arrest assay for the evaluation of parameters affecting intrastrand tetraplex formation. J. Biol. Chem. *271*, 20958–20964.

Wilkins, M.H.F., Stokes, A.R., and Wilson, H.R. (1953). Molecular Structure of Nucleic Acids: Molecular Structure of Deoxypentose Nucleic Acids. Nature *171*, 738–740.

Wilkinson, K.A., Merino, E.J., and Weeks, K.M. (2006). Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. Nat. Protoc. *1*, 1610–1616.

Williamson, J.R., Raghuraman, M.K., and Cech, T.R. (1989). Monovalent cation-induced structure of telomeric DNA: The G-quartet model. Cell *59*, 871–880.

Wolfe, A.L., Singh, K., Zhong, Y., Drewe, P., Rajasekhar, V.K., Sanghvi, V.R., Mavrakis, K.J., Jiang, M., Roderick, J.E., Van der Meulen, J., et al. (2014). RNA G-quadruplexes cause eIF4A-dependent oncogene translation in cancer. Nature *513*, 65–70.

Wong, H.M., and Huppert, J.L. (2009). Stable G-quadruplexes are found outside nucleosome-bound regions. Mol. Biosyst. *5*, 1713–1719.

Woodford, K.J., Howell, R.M., and Usdin, K. (1994). A novel K+-dependent DNA synthesis arrest site in a commonly occurring sequence motif in eukaryotes. J. Biol. Chem. *269*, 27029–27035.

Workman, J.L., and Kingston, R.E. (1992). Nucleosome core displacement in vitro via a metastable transcription factor-nucleosome complex. Science *258*, 1780–1784.

Wu, T., Lyu, R., You, Q., and He, C. (2020). Kethoxal-assisted single-stranded DNA sequencing captures global transcription dynamics and enhancer activity in situ. Nat. Methods *17*, 515–523.

Xiao, S., Zhang, J.Y., Wu, J., Wu, R.Y., Xia, Y., Zheng, K.W., Hao, Y.H., Zhou, X., and Tan, Z. (2014). Formation of DNA:RNA hybrid G-quadruplexes of two G-quartet layers in transcription: Expansion of the prevalence and diversity of G-quadruplexes in genomes. Angew. Chemie - Int. Ed. *53*, 13110–13114.

Xu, H., Di Antonio, M., McKinney, S., Mathew, V., Ho, B., O'Neil, N.J., Santos, N. Dos, Silvester, J., Wei, V., Garcia, J., et al. (2017). CX-5461 is a DNA G-quadruplex stabilizer with selective lethality in BRCA1/2 deficient tumours. Nat. Commun. *8*, 14432.

Xu, Y., Kaminaga, K., and Komiyama, M. (2008). G-quadruplex formation by human telomeric repeats-containing RNA in Na + solution. J. Am. Chem. Soc. *130*, 11179–11184.

Xu, Y., Suzuki, Y., and Komiyama, M. (2009). Click chemistry for the identification of g-quadruplex structures: Discovery of a DNA-RNA g-quadruplex. Angew. Chemie - Int. Ed. *48*,

3281–3284.

Xu, Y., Suzuki, Y., Ito, K., and Komiyama, M. (2010). Telomeric repeat-containing RNA structure in living cells. Proc. Natl. Acad. Sci. U. S. A. *107*, 14579–14584.

Xue, Y., Wong, J., Moreno, G.T., Young, M.K., Côté, J., and Wang, W. (1998). NURD, a novel complex with both ATP-dependent chromatin-remodeling and histone deacetylase activities. Mol. Cell *2*, 851–861.

Yamada, T., Yamaguchi, Y., Inukai, N., Okamoto, S., Mura, T., and Handa, H. (2006). P-TEFb-mediated phosphorylation of hSpt5 C-terminal repeats is critical for processive transcription elongation. Mol. Cell *21*, 227–237.

Yamaguchi, Y., Takagi, T., Wada, T., Yano, K., Furuya, A., Sugimoto, S., Hasegawa, J., and Handa, H. (1999). NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. Cell *97*, 41–51.

Yan, F., Powell, D.R., Curtis, D.J., and Wong, N.C. (2020). From reads to insight: A hitchhiker's guide to ATAC-seq data analysis. Genome Biol. *21*, 1–16.

Yang, J.G., Madrid, T.S., Sevastopoulos, E., and Narlikar, G.J. (2006). The chromatin-remodeling enzyme ACF is an ATP-dependent DNA length sensor that regulates nucleosome spacing. Nat. Struct. Mol. Biol. *13*, 1078–1083.

Ying, L., Green, J.J., Li, H., Klenerman, D., and Balasubramanian, S. (2003). Studies on the structure and dynamics of the human telomeric G quadruplex by single-molecule fluorescence resonance energy transfer. Proc. Natl. Acad. Sci. U. S. A. *100*, 14629–14634.

Yu, Z., Schonhoft, J.D., Dhakal, S., Bajracharya, R., Hegde, R., Basu, S., and Mao, H. (2009). ILPR G-quadruplexes formed in seconds demonstrate high mechanical stabilities. J. Am. Chem. Soc. *131*, 1876–1882.

Zahler, A.M., Williamson, J.R., Cech, T.R., and Prescott, D.M. (1991). Inhibition of telomerase by G-quartet DNA structures. Nature *350*, 718–720.

Zaret, K.S., and Carroll, J.S. (2011). Pioneer transcription factors: Establishing competence for gene expression. Genes Dev. *25*, 2227–2241.

Zhang, J. yu, Xia, Y., Hao, Y. hua, and Tan, Z. (2020). DNA:RNA hybrid G-quadruplex formation upstream of transcription start site. Sci. Rep. *10*, 1–6.

Zhang, M.L., Tong, X.J., Fu, X.H., Zhou, B.O., Wang, J., Liao, X.H., Li, Q.J., Shen, N., Ding, J., and Zhou, J.Q. (2010a). Yeast telomerase subunit Est1p has guanine quadruplex-promoting activity that is required for telomere elongation. Nat. Struct. Mol. Biol. *17*, 202–209.

Zhang, S., Sun, H., Wang, L., Liu, Y., Chen, H., Li, Q., Guan, A., Liu, M., and Tang, Y. (2018). Real-time monitoring of DNA G-quadruplexes in living cells with a small-molecule

fluorescent probe. Nucleic Acids Res. *46*, 7522–7532.

Zhang, Z., Dai, J., Veliath, E., Jones, R.A., and Yang, D. (2010b). Structure of a two-G-tetrad intramolecular G-quadruplex formed by variant human telomeric sequence in K+ solution: Insights into the interconversion of human telomeric G-quadruplex structures. Nucleic Acids Res. *38*, 1009–1021.

Zhao, K., Jia, Y., Peng, J., Pang, C., Zhang, T., Han, W., Jiang, J., Lu, X., Zhu, J., and Qian, Y. (2019). Anacardic acid inhibits RANKL-induced osteoclastogenesis in vitro and prevents ovariectomy-induced bone loss in vivo. FASEB J. *33*, 9100–9115.

Zheng, K., Zhang, J., He, Y., Gong, J., Wen, C., Chen, J., Hao, Y., Zhao, Y., and Tan, Z. (2020). Detection of genomic G-quadruplexes in living cells using a small artificial protein. Nucleic Acids Res. 1–15.

Zheng, K.W., Xiao, S., Liu, J.Q., Zhang, J.Y., Hao, Y.H., and Tan, Z. (2013). Co-transcriptional formation of DNA: RNA hybrid G-quadruplex and potential function as constitutional cis element for transcription control. Nucleic Acids Res. *41*, 5533–5541.

Zyner, K.G., Mulhearn, D.S., Adhikari, S., Cuesta, S.M., Di Antonio, M., Erard, N., Hannon, G.J., Tannahill, D., and Balasubramanian, S. (2019). Genetic interactions of G-quadruplexes in humans. Elife *8*, 1–40.